LARGE-SIGNAL MODEL DEVELOPMENT AND HIGH EFFICIENCY POWER AMPLIFIER DESIGN IN CMOS TECHNOLOGY FOR MILLIMETER-WAVE APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

by

Navin Mallavarpu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2012

LARGE-SIGNAL MODEL DEVELOPMENT AND HIGH EFFICIENCY POWER AMPLIFIER DESIGN IN CMOS TECHNOLOGY FOR MILLIMETER-WAVE APPLICATIONS

Approved by:

Dr. Manos Tentzeris, Co-Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Saibal Mukhopadhyay
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Debasis Dawn, Co-Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Kevin Kornegay
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. David Hertling
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Brent Wagner
Georgia Tech Research Institute
Georgia Institute of Technology

Date Approved: April 12, 2012

ACKNOWLEDGEMENTS

I wish to thank the professors serving on my Ph.D. committee for taking the time to review my research work and dissertation. I would also like to thank my current advisors and former advisor for providing guidance for my Ph.D. research. The funding provided by faculty, the ECE Department and the Georgia Tech Research Institute were all necessary. I would also like to acknowledge my parents and brother for their continued support and encouragement.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	xii
CHAPTER	
1 INTRODUCTION	1
2 MILLIMETER-WAVE CMOS DEVICE CHARACTERIZATION	5
2.1 CMOS Device Structure and Processes	5
2.2 Device Measurements and Calibration	9
2.3 Measurement De-embedding	14
2.4 Preliminary Modeling Approach – BSIM3 Macromodel	17
3 TEMPERATURE-DEPENDENT CMOS LARGE SIGNAL MODEL FOR MILLIMETER-WAVE APPLICATIONS	25
3.1 Basic Model Development	25
3.2 CMOS Short Channel Device Physics	28
3.2.1 Velocity Saturation	29
3.2.2 Channel Length Modulation	30
3.2.3 Drain-Induced Barrier Lowering	31
3.3 Nonlinear Drain Current Source	32
3.4 Circuit Model Extraction for Millimeter-wave Applications	41
3.5 Implementation of Temperature-Dependency and Size Scalability	49

4 60 GHZ CMOS POWER AMPLIFIER DESIGN 58

5 24 GHz INVERSE CLASS F POWER AMPLIFIER DESIGN IN 130nm CMOS PROCESS 72

5.1 Preliminary Considerations – Classes of Operation 72

5.2 Use of BiCMOS 130nm 8HP Process 79

5.3 Design Approach 83

5.4 Layout Approach 93

5.5 Measured and Simulated PA Characteristics 95

6 SUMMARY AND CONCLUSIONS 100

6.1 Summary 100

6.2 Conclusions 100

6.3 Contributions 102

6.4 Suggestions for Future Research 104

6.5 Publications Resulting from this Work 104

REFERENCES 105

VITA 109

LIST OF TABLES

Page

Table 2.1: Summary of device characteristics for 90nm ST Microelectronics CMOS process. 9

Table 3.1: Description of parameters in Ids model equations. 34

Table 3.2: Numerical values of drain current equation for 40μm, 90nm CMOS device. 40

Table 3.3: Summary of parameters used for incorporation of temperature dependency into drain current equation. 53

Table 4.1: Performance summary of three-stage 60GHz CMOS Power Amplifier. 67

Table 4.2: Performance summary of four-stage 60GHz CMOS Power Amplifier. 71

Table 5.1: Summary of tradeoffs between classes of operation. 78

LIST OF FIGURES

Page

Figure 1.1: Applications enabled by millimeter-wave CMOS-based systems on chip. 1

Figure 1.2: Example of a fully-integrated front-end chip used in millimeter-wave applications. 2

Figure 1.3: Die photo of integrated transmitter/receiver CMOS chip used in millimeter-wave applications. 3

Figure 2.1: MOSFET cross-sectional view. 5

Figure 2.2: FET structure layout view. 6

Figure 2.3: Load-Pull and large-signal power measurement setup. 10

Figure 2.4: ISS calibration structures. 12

Figure 2.5: Die photo of a device test structure with GSG pads and feeding lines. 14

Figure 2.6: (a) Open Structure, (b) Short Structure, (c) Device Test Structure. DC Bias pads not shown. 15

Figure 2.7: Test structure de-embedding procedure used for active devices. 16

Figure 2.8: Schematic of modified BSIM3-based circuit model. 18

Figure 2.9: Device capacitances versus temperature. 19

Figure 2.10: Device capacitances versus device size. 19

Figure 2.11: Gate resistance versus temperature. 20

Figure 2.12: S21 magnitude versus device size. 21

Figure 2.13: S21 magnitude versus temperature. 22

Figure 2.14: Example of a harmonic balance convergence problem encountered with the BSIM3-macromodel vs. the expected power amplifier large signal power characteristic. 23

Figure 3.1: Flowchart showing model extraction procedure. 26

Figure 3.2: Schematic of circuit model with bulk tied to source. 27

Figure 3.3: Drift velocity versus lateral electric field 29

Figure 3.4: Channel length reduction due to increasing Vds in short-channel device. 31

Figure 3.5: Representative Ids vs. Vds curve for a fixed Vgs value in strong inversion. 33

Figure 3.6: Representative Ids vs. Vgs curve for a fixed Vds value. 33

Figure 3.7: Heaviside step function for various parameter values. 35

Figure 3.8: Ids vs. Vds (0-1.5V) for Vgs (0 to 1.0V) for an 80x1 μm, 90nm NMOS device. 36

Figure 3.9: Ids vs. Vgs (0-1.5V) for Vgs (0 to 1.0V) for an 80x1 μm, 90nm NMOS device. 36

Figure 3.10: gm vs. Vgs (Vds=0.6, 0.9 and 1.2V) for 80μm device. 37

Figure 3.11: Ids vs. Vds for Vgs=0.1 and 0.2V. 39

Figure 3.12: Agilent ADS implementation of basic version of circuit model with the SDD- based drain current generator. 40

Figure 3.13: Model extraction flowchart 41

Figure 3.14: Cgs vs. Vgs for varying Vds. 44

Figure 3.15: Cgd vs. Vds for Vgs = 0.4 and 0.6V. 45

Figure 3.16: Model capacitances versus device size. 47

Figure 3.17: Model Rg, Rd and Rs as function of device size. 47

Figure 3.18: Model gate inductance as function of device size. 48

Figure 3.19: Gate resistance vs. temperature for a 40μm device. 49

Figure 3.20: Ids vs. temperature and Vgs for Vds= 1.0V for a 40x1 μm, 90nm NMOS device. 50

Figure 3.21: Ids vs. Vgs for various temperatures (0 to 80 degrees Celsius) for a 40x1 μm, 90nm NMOS device. 50

Figure 3.22: Third-order polynomial as a function of temperature used in temperature- dependent implementation of Ids. 52

Figure 3.23: Mobility vs. electric field characteristic in a FET with scattering mechanisms which form the characteristic. 54

Figure 3.24: Simulated vs. measured S21 vs. temperature for 40x1 μm device for Vds=1.0V, Vgs=0.6V. 55

Figure 3.25: Simulated vs. measured magnitude and phase of S21, S11 and S22 at fixed temperature, 27C. 56

Figure 3.26: Measured and simulated large signal power sweeps for 40μm and 80μm at room temperature, 27 degrees. 57

Figure 4.1: Example of load pull simulation bench in Agilent ADS. 59

Figure 4.2: Power and PAE contours for a 40μm NMOS transistor at 60 GHz based on developed large signal model. 60

Figure 4.3: Power and PAE contours for an 80μm NMOS transistor at 60 GHz based on developed large signal model. 60

Figure 4.4: Optimum load and source impedances for 40, 80 and 160μm devices at 60GHz. 60

Figure 4.5: DC Loadline for maximum output power for a FET with Class A quiescent bias. 62

Figure 4.6: Power Amplifier Design Flow. 64

Figure 4.7: Schematic of three-stage 60GHz CMOS power amplifier. 65

Figure 4.8: Measured and simulated S-Parameters of three-stage 60GHz CMOS PA. 66

Figure 4.9: Measured and simulated power performance of three-stage 60GHz CMOS PA over temperature. 66

Figure 4.10: Comparison between large signal power performance of three stage PA with 40 and 60μm output stages. 67

Figure 4.11: Schematic of four-stage power amplifier. 68

Figure 4.12: Die Photo of four-stage power amplifier. 69

Figure 4.13: Measured and simulated S-Parameters of four-stage power amplifier. 69

Figure 4.14: Large signal power performance of four-stage power amplifier. 70

Figure 4.15: Power added efficiency of four-stage power amplifier at 60GHz. 70

Figure 4.16. Measured and simulated drain current versus input power of four-stage power amplifier. 71

Figure 5.1: Load-lines for non switching mode classes of power amplifier operation. 73

Figure 5.2: Drain current waveforms for Class A, Class B and Class C Modes for the same input drive. 74

Figure 5.3: Ideal drain current and voltage waveforms for Class F and F Inverse Power Amplifiers. 77

Figure 5.4: Ideal tuned Class F PA output matching network. 78

Figure 5.5: Stack for IBM 8HP process 80

Figure 5.6: (a) DC I-V curves and (b) transconductance curves. 81

Figure 5.7: S-Parameters for a 40µm device. 81

Figure 5.8: Load Pull simulation results for 80µm model with harmonic load impedances set to Inverse Class F values. 83

Figure 5.9: Simulated Class AB current and voltage waveforms using 80µm device model. 84

Figure 5.10: Simulated Inverse Class F current and voltage waveforms using 80µm device model. 84

Figure 5.11: Simulated Class F current and voltage waveforms using 80µm device model. 85

Figure 5.12: Smith Chart with load and source targets for 24GHz Inverse Class F PA. 85

Figure 5.13: Overview of output matching network for 24GHz Inverse Class F PA. 86

Figure 5.14: Output matching network with ideal transmission lines for 24GHz Inverse Class F PA. 88

Figure 5.15: 24GHz Inverse Class F schematic. 89

Figure 5.16: Smith Chart with realized output matching network impedances – ADS simulation. 90

Figure 5.17: Drain current and voltage waveforms for Inverse Class F PA. 90

Figure 5.18: ADS schematic of PA. 91

Figure 5.19: K-factor simulation of 24GHz Inverse Class F PA. 92

Figure 5.20: Cadence layout of 24GHz Inverse Class F PA. 93

Figure 5.21: Die photo of 24GHz Inverse Class F PA. 95

Figure 5.22: Measured (Blue) and Simulated (Red) S21 Magnitude of 24GHz Inverse Class F PA. 96

Figure 5.23: Measured (Blue) and Simulated (Red) S11 Magnitude of 24GHz Inverse Class F PA. 96

Figure 5.24: Measured (Blue) and Simulated (Red) S22 Magnitude of 24GHz Inverse Class F PA. 97

Figure 5.25: Measured (Blue) and Simulated (Red) S12 Magnitude of 24GHz Inverse Class F PA. 97

Figure 5.26: Measured (Dots) and Simulated (Solid Curve) large signal power sweep of 24GHz Inverse Class F PA. 98

Figure 5.27: Measured (Dots) and Simulated (Solid Curve) power-added efficiency of 24GHz Inverse Class F PA. 98

SUMMARY

This dissertation presents a novel large signal modeling approach which can be used to accurately model CMOS transistors used in millimeter-wave CMOS power amplifiers. The large signal model presented in this work is classified as an empirical compact device model which incorporates temperature-dependency and device periphery scaling. These added features allow for efficient design of multi-stage CMOS power amplifiers by virtue of the process-scalability. Prior to the presentation of the details of the model development, background is given regarding the 90nm CMOS process, device test structures, de-embedding methods and device measurements, all of which are necessary preliminary steps for any device modeling methodology. Following discussion of model development, the design of multi-stage 60GHz Class AB CMOS power amplifiers using the developed model is shown, providing further model validation. The body of research concludes with an investigation into designing a CMOS power amplifier operating at frequencies close to the millimeter-wave range with a potentially higher-efficiency class of power amplifier operation. Specifically, a 24GHz 130nm CMOS Inverse Class F power amplifier is simulated using a modified version of the device model, fabricated and compared with simulations. This further demonstrates the robustness of this device modeling method.

CHAPTER 1

INTRODUCTION

Silicon technology, encompassing both field-effect transistor (FET) based complementary metal oxide semiconductor (CMOS) and heterojunction bipolar transistor (HBT) based Silicon Germanium (SiGe) processes, has emerged as a significant enabler of fully-integrated RF, Microwave and Millimeter-wave systems-on-chip. Some of the advantages of using Silicon CMOS transistor technology for these applications include the high degree of integration that is possible on a single chip, the low cost of the Silicon die and the multiple layers of metallization available in the CMOS processes. Processes with thick top metal layers can be used for RF through millimeter wave applications by allowing for microstrip and coplanar transmission line implementation. Since most digital functions are already implemented in standard CMOS processes, the reduction in the gate length of standard CMOS processes has made possible the integration of the entire RF front ends with digital control circuitry on a single chip used for transmitting and receiving wireless signals. At millimeter-wave frequencies, including 60 GHz, short-range communication between devices with wireless capability is one application of such integrated chips. Figure 1.1 illustrates some examples of these applications and the type of data rates required for each.

Figure 1.1 Applications enabled by millimeter-wave CMOS-based systems on chip.

An example of a fully-integrated millimeter-wave front end complete with digital control and signal processing in addition to the RF transmitter and receiver chains is shown in Figure 1.2. A fully body of literature, encompassed in part by the research work presented in [1]-[11], demonstrates the advancements made in developing Silicon-based millimeter-wave transmitters and receivers fully integrated on a single chip.

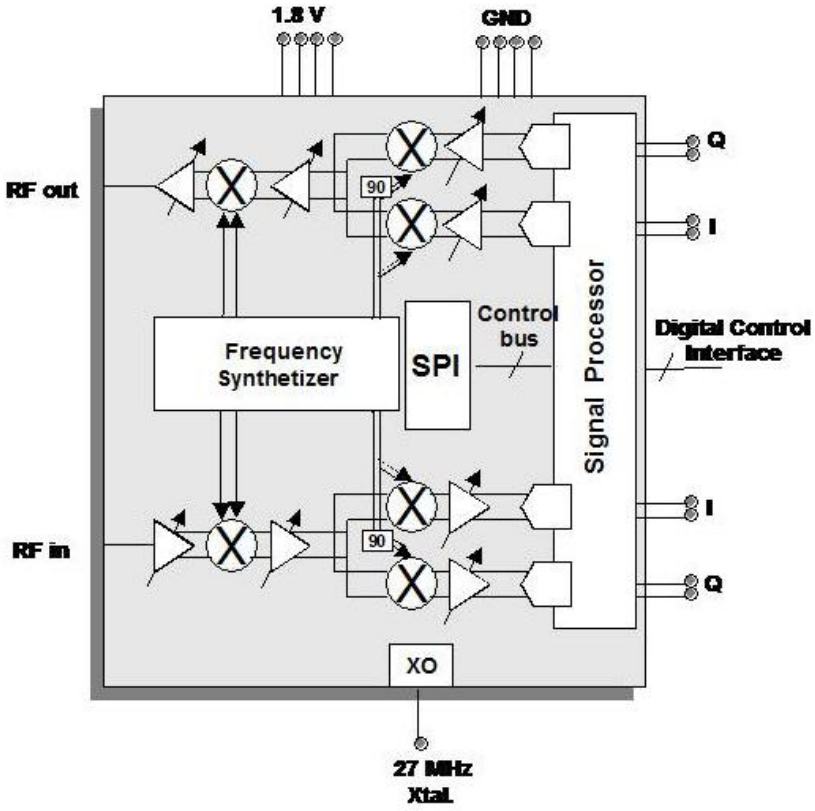


Figure 1.2 Example of a fully-integrated front-end chip used in millimeter-wave applications.

Using 60GHz as the frequency of transmission and reception is beneficial due to the fact that at 60GHz a very high percentage (98%) of the transmitted signal is absorbed by atmospheric oxygen [12]. While this high rate of absorption is highly problematic for any sort of long range transmission and reception, it can actually be beneficial for short-

range applications such as those shown in Figure 1.1, in which the communications devices are within a couple of meters of each other. On the other hand, the benefit of this oxygen absorption is that frequency re-use is very practical for different sets of devices that exist within a small area.

A die photo of a complete, integrated chip with transmitter and receiver chains is shown in Figure 1.3.

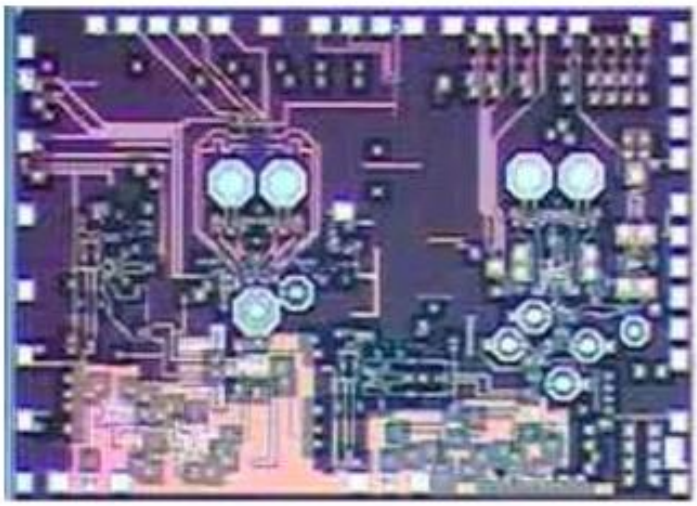


Figure 1.3. Die Photo of integrated transmitter/receiver CMOS chip used in millimeter-wave applications.

The last part of the transmitter chain in such chips is the power amplifier which is a highly critical component that needs to be designed properly and accurately to ensure first-pass design success. This is necessary due to the high cost of running successive tape-outs. Accurate device models are needed, in both the small and large signal regimes, to ensure successful designs which meet performance targets. The power amplifier must be designed to provide sufficient output power on the transmitter side, as efficiently as possible, using the DC power and input RF power. Furthermore, the overall gain of the

power amplifier must be high enough so that the low power level input from the previous part of the transmitter chain can be amplified to produce the desired saturated output power. Examples of CMOS power amplifiers designed at 60GHz are found in [13]-[16].

This dissertation focuses on the development of a novel device modeling approach for CMOS-based field-effect transistors used in millimeter-wave CMOS power amplifier development. The incorporation of temperature dependency and device size scalability enables the efficient design of multi-stage CMOS power amplifiers by accounting for process variation and variable operating conditions. Accurate modeling of the power amplifiers across a range of temperatures and not just at room temperature is critical because it allows the designer to decide whether the gain, power and efficiency of the power amplifier vary significantly enough across a temperature range to warrant incorporating biasing techniques that make the amplifier performance stable across that range of temperatures. Following the development of the modeling approach, examples of designs developed and fabricated using this model are shown at 60GHz. Following this, a design is implemented using this model at 24GHz, the goal being to investigate the use of a potentially higher efficiency mode of power amplifier operation, that could in the future, be applicable at 60GHz if given a process with a short-enough gate length.

CHAPTER 2

MILLIMETER-WAVE CMOS DEVICE CHARACTERIZATION

2.1 CMOS Device Structure and Processes

The CMOS processes used in both the model development and power amplifier design for millimeter-wave applications (through 65 GHz) feature short gate-lengths on Silicon substrate. CMOS processes with gate lengths of 90 and 130 nm have been made available for the research conducted in this work. A cross-sectional view of a standard NMOS field-effect transistor is shown in Figure 2.1.

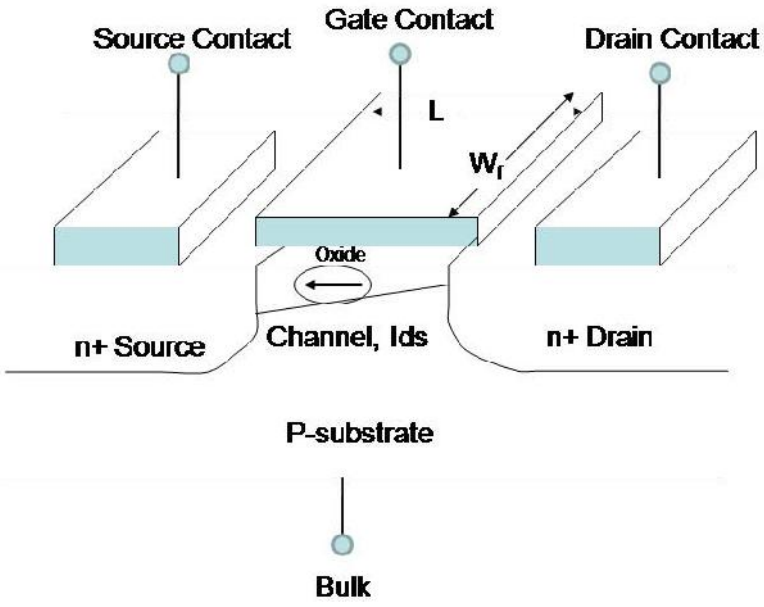


Figure 2.1. MOSFET cross-sectional view

Since in a particular process, the channel length, L, is fixed, it is necessary to adjust the width parameter, W, of the device for greater current drive.

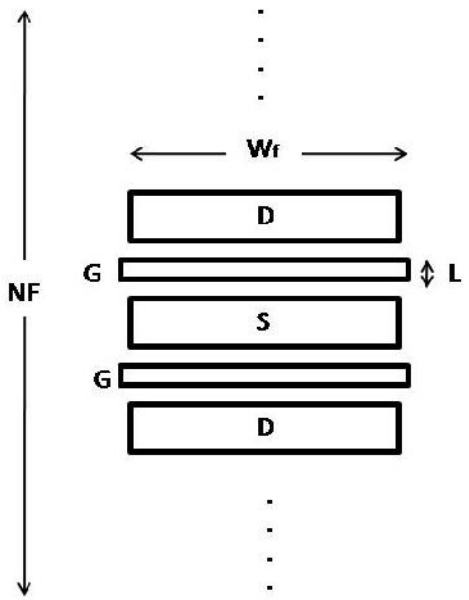


Figure 2.2. FET structure layout view

The width parameter, W, is the total device width or periphery and is a product of the number of gate fingers, NF, and the width of each gate finger as follows:

$$W = NF \cdot W_{Finger} \quad (2.1)$$

Any time the variable W is used from here forward, it refers to the total device width or periphery. Larger low frequency gain is obtained with larger periphery devices however there is also faster frequency roll-off with increasing periphery due to larger device capacitances. In this work, NMOS transistors were designed, laid-out and tested, with device peripheries ranging from 20µm to 80µm in increments of 20µm, followed by two 80µm devices combined in parallel, yielding 160µm total periphery. In subsequent chapters, large-signal power simulations of models based on these devices, will show the increasing saturated output power with increasing device periphery.

Since we are concerned with millimeter-wave device performance, it is necessary to calculate the Ft and Fmax values for each device fabricated and ensure that they are high enough for 60 GHz power amplifier development. Both of these parameters increase with decreasing gate length, hence the need for using short-gate length devices, generally with gate lengths around 100nm or smaller, for millimeter-wave applications. The Ft of a transistor can be determined from the frequency at which the magnitude of the short circuit current gain of the transistor, h_{21}, is equal to unity (or equivalently 0 dB on a logarithmic plot). If a device model is available, the Ft of the device can be approximated as:

$$f_T = \frac{gm}{Cgs} \qquad (2.2)$$

In this equation, gm is the device transconductance at a particular bias point and Cgs is the gate-to-source capacitance. Likewise, the fmax is the frequency at which the maximum available gain of the transistor falls to unity (or equivalently 0 dB on a logarithmic plot).

The relation between f_{max} and f_T yields an important result that is critical to understand when trying to determine what should be the width of each gate finger. In other words, it can determine the limit of acceptable gate finger widths. First, the polysilicon gate resistance of a gate finger should be minimized. It is obvious from (2.3) that since R_{poly} and n are process parameters and L is fixed that W is the only parameter that can be varied when creating new device structures.

The gate resistance equation is shown below:

$$R_g = \frac{R_{poly} \cdot W_{finger}}{3 \cdot n^2 \cdot L} \qquad (2.3)$$

Next it is necessary to examine the aforementioned relationship between f_{max} and f_T in (2.4):

$$f_{max} = \frac{f_T}{2 \cdot \sqrt{R_g \cdot \left(gm \cdot \frac{Cgd}{Cgs}\right) + (R_g + r_{ch} + R_s) \cdot g_{ds}}} \qquad (2.4)$$

It is evident that there is a strong dependence on R_g and that minimizing this will improve the f_{max} value for the device with respect to f_T. Of course that is not to say that R_g can be made as small as possible (by making W_{finger} as small as possible), for this would reduce the device gain considerably and require a very large number of gate fingers in parallel to have a large enough device periphery. So given this tradeoff, it is found that a finger width of around 1µm to at most 2µm is acceptable. Given this, to achieve the previously mentioned peripheries of 20µm, 40µm, 80µm and so on, device test structures were designed and fabricated with 20 gate fingers each with 1µm gate width and similarly for the larger peripheries. When requiring larger periphery devices capable of producing higher saturated output power (Psat), there is an upper limit to take into consideration. Generally, 80 fingers is the limit because greater than this value, the resistive losses associated with the gate, source and drain connecting structures starts to get too high, especially at millimeter-wave frequencies. Source degeneration also increases unacceptably. So for larger periphery devices, two smaller cells are combined. All of these devices were designed and fabricated in the 90nm ST Microelectronics CMOS process, the device gate length being 100nm. In Table 2.1, a summary of the

device characteristics is given. Imax is the maximum drain current of the transistor, Ropt is calculated from the loadline value for maximum power output of the device and Copt is the reactance to be used in parallel with Ropt, for maximum power output matching. Note that these are normalized quantities and can be used for design purposes with transistors of varying periphery.

Table 2.1. Summary of device characteristics for 90nm ST Microelectronics CMOS process.

Gate Finger Width of Transistors	1-2 μm
Range of Total Device Widths	20-160 μm
Ft (Vds=0.9V,Vgs=0.6V)	150 GHz
Fmax (Vds=0.9V,Vgs=0.6V)	170 GHz
Imax	0.7A/mm
Ropt	2.1ohm*mm
Copt	-1.16 pF/mm

2.2 Device Measurements and Calibration

To obtain all of the device measurements necessary for device model development and subsequent power amplifier design requires extensive measurements (both DC and RF). This in turn requires very careful and frequent measurement setup calibration, especially at millimeter-wave frequencies, where measurement system drift over the span of a few hours is one problem. Even following accurate measurements, precise de-embedding of the ground-signal-ground (GSG) pad feeding structures to the device under test (DUT) is needed. Without any of these, it will not be possible to develop accurate

device models and will require more guess-work to achieve first pass design success. The measurements include DC, S-Parameter, Load-pull and large-signal power measurements and all are “on-wafer” and are performed using ground-signal-ground probes with the device die held in place by vacuum on a wafer chuck. A block diagram of the measurement setup for a DUT (transistor in the case of load-pull and either a transistor or an amplifier for large-signal power sweep) is shown in Figure 2.3.

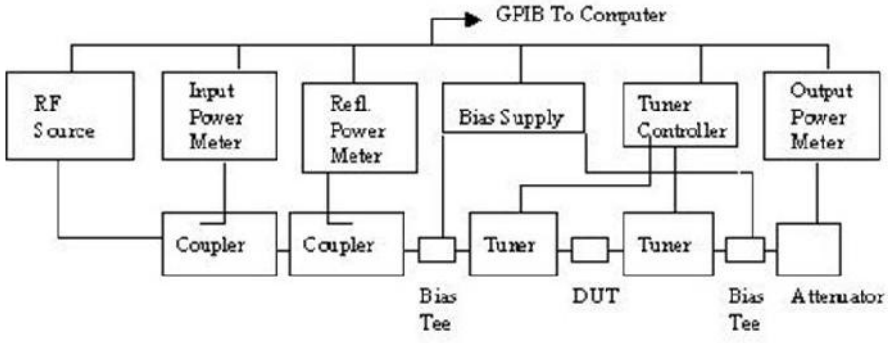


Figure 2.3. Load-Pull and Large-Signal Power Measurement setup

For doing small-signal measurements on wafer, the primary difference is the substitution of a Vector Network Analyzer (VNA) for the RF power source. The load and source tuners are not needed nor are the power meters. Small-signal measurements require RF input power levels of generally less than -30dBm. Prior to doing any of these measurements, however, a probe calibration is required. The probe calibration refers to a process of calibrating out the losses (such as cable and connector losses) between the RF source and the GSG probe tips, which make contact with the GSG pad structures of the DUT. This is a critical step for performing measurements at millimeter-wave frequencies. There are various methods for doing this, however some are better at millimeter-wave

frequencies. All of the methods however, utilize an impedance standard substrate, otherwise known as an ISS. All of the on-wafer measurements were conducted using a semi-automatic probe station provided by Cascade Microtech. The probes used are air-coplanar (ACP) GSG probes also provided by Cascade. Separate DC probes are used for biasing the gate and drain of the DUT. The primary probe calibration methods include Short-Open-Load-Thru (SOLT), Thru-Reflect-Line (TRL), Line-Reflect-Match (LRM) and a variant of the latter, Line-Reflect-Reflect-Match (LRRM). The pros and cons of each of these will be explained and the calibration method with the most advantages (or fewest disadvantages) for millimeter-wave measurements will be determined. The SOLT calibration method is on every VNA and is probably the most commonly used, however it has some inherent disadvantages. Specifically, some drawbacks include the need for perfect probe placement, the short, load and open parasitics, and overall calibration inconsistency. The TRL approach is not the most accurate, and requires multiple transmission line standards, thereby requiring multiple probe spacings. The LRRM (and LRM) approach requires only one transmission line standard (for the thru line), has broadband calibration accuracy, and is available in the automated measurement and calibration software, such as WinCal provided by Cascade. Given the pros and cons of each of these methods, the LRRM probe calibration method was used to accurately measure small and large signal device operation at the GSG pad reference planes of the device test structures as well as for the power amplifiers subsequently developed. For illustration purposes, diagrams of the various ISS calibration standards are shown in Figure 2.4. General theory of load and source pull measurements and calibration is based on the fundamentals presented in [17] and [18].

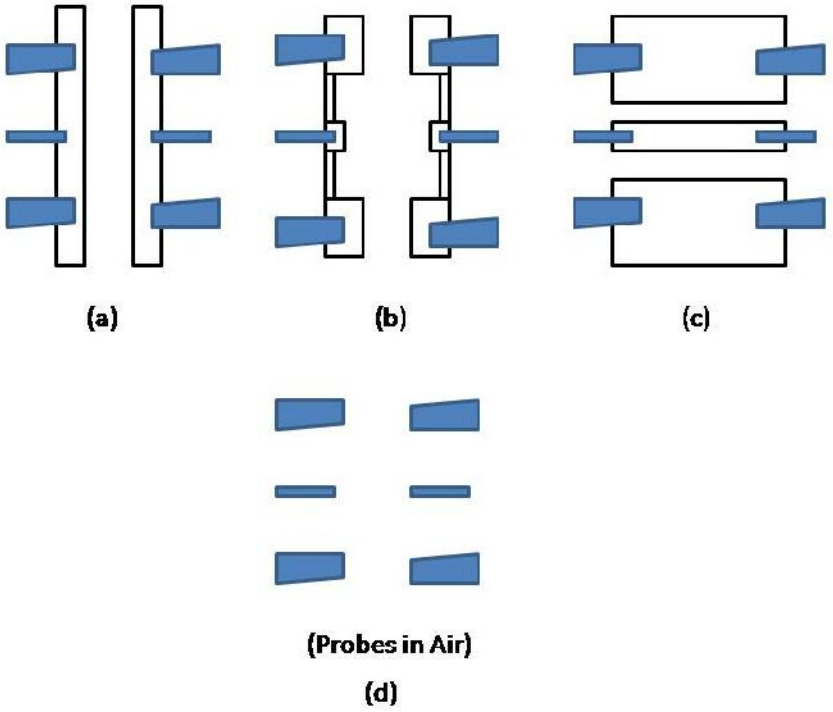


Figure 2.4 ISS Calibration Structures – (a) Short (b) Load (c) Thru or Line (d) Open. GSG probe tips are shown for illustration purposes.

Following accurate probe calibration through 65GHz, small signal measurements were performed to obtain S-parameters of the device test structures (ranging in total periphery from 20 to 160 μm), using a calibrated vector network analyzer. These measurements obviously need DC probes as well, so DC I-V measurements are performed first to obtain DC characteristics and to determine whether the device is functional. Following this, large-signal power testing of the devices can be performed. A brief overview of the power bench calibration and measurement procedures will be given. First, all of the components in the power measurement system were disconnected and their S-parameters were measured using a calibrated Vector Network Analyzer (VNA). These S-parameters

were stored in blocks for each of the individual components, to be used later for determining simulated power gain values, which would then be compared to the power measurement results. The Maury Automated Tuner System (ATS) software was used to store these S-parameters during the calibration procedure and to control the actual load and source pull measurements during the actual power testing. When checking the calibration of the power bench using a thru line, it was first necessary to make sure that the predicted Gt (determined by de-embedding S-parameters of the calibrated components) and the measured Gt for the thru line agreed satisfactorily. Checking the calibration involved a series of load and source pulls to determine the limits of the measurements for the power bench, and determine which components of the power bench needed to have S-parameters re-measured. An iterative process then ensued if it was deemed necessary, until satisfactory calibration results were obtained. For the actual power measurements, the Maury ATS software provided Smith Chart displays for the load and source tuners allowing for easy selection of source and load impedances presented by the respective tuners. The basic methodology was to first determine a small region of the Smith Chart for which the source reflection coefficient could be minimized, keeping the load impedance fixed at the center of the Smith chart. Then using this source impedance, the load impedance was varied to determine an optimal point for output power, gain and power-added efficiency. The process involved going back and forth between source and load tuners to achieve optimal tuning.

2.3 Measurement De-embedding

In Figure 2.5, a die photo of a typical mm-wave transistor test structure is shown. The actual transistor core is at the very center of the two feeding microstrip lines extending from the signal pads on each side of the structure. The two sets of pads at the top and bottom of the test structure are ground pads. It is evident that the device measurement process cannot end with the measurements themselves and that the procedure of de-embedding is necessary to obtain device characteristics at the device plane. There are various methods of doing this, however the method employed in this work is open-short de-embedding. This applies mainly to the transistor and amplifier S-Parameter measurements.

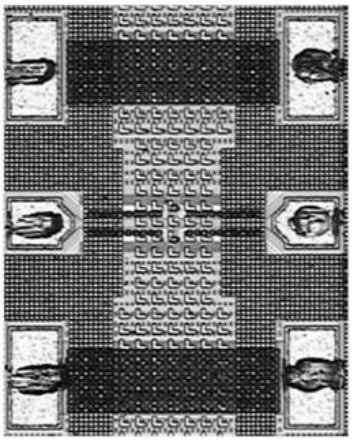


Figure 2.5. Die photo of a device test structure with GSG pads and feeding lines.

In this approach, the measured S-Parameters are first measured for the DUT test structure, an open test structure and a short test structure. Diagrams of each of these are shown for illustration purposes in Figure 2.6.

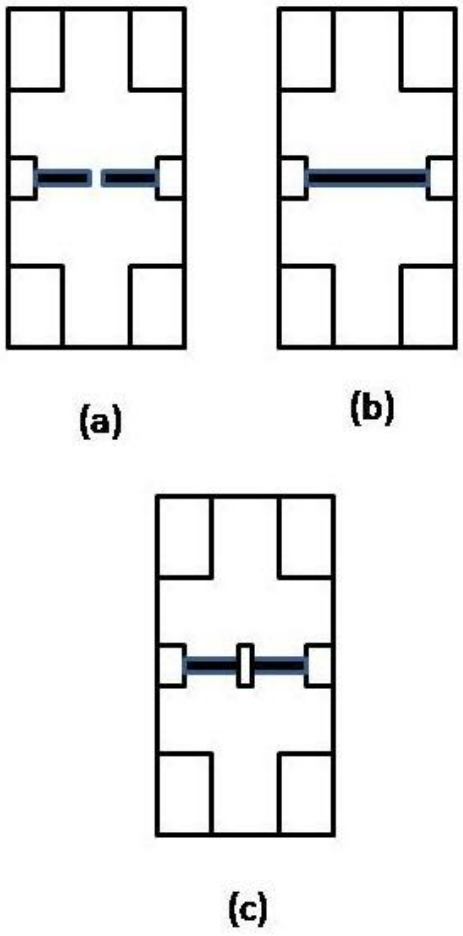


Figure 2.6. (a) Open Structure, (b) Short Structure, (c) Device Test Structure.
DC Bias pads not shown.

Once the S-Parameters of each of these test structures are measured, they are stored in what is called an S-Matrix. This results in the matrices $\mathbf{S_{open}}$, , $\mathbf{S_{short}}$ and $\mathbf{S_{test\_str}}$. These matrices are then converted to Y-parameters. There are several steps necessary after this

to arrive at the de-embedded S-Parameters of the DUT. The entire procedure is summarized in the flowchart in Figure 2.7.

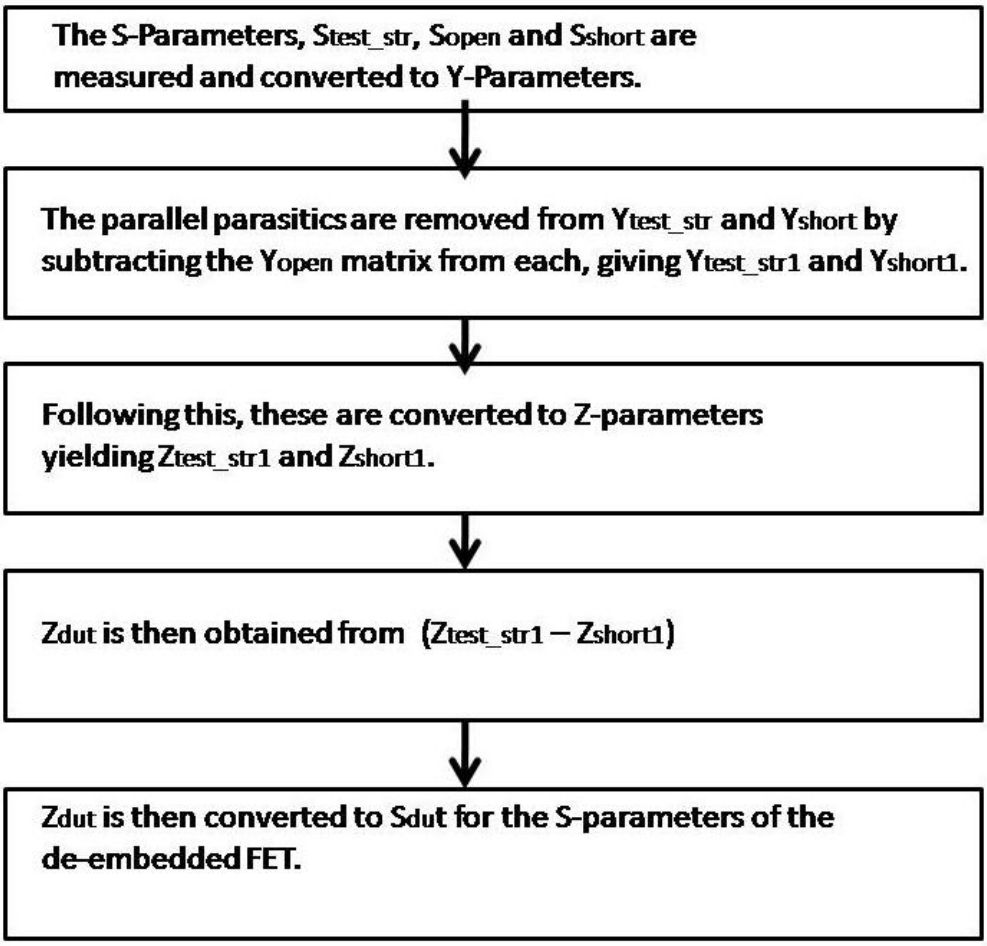


Figure 2.7. Test Structure de-embedding procedure used for active devices.

2.4 Preliminary Modeling Approach – BSIM3-based Macromodel

The first step in any power amplifier design involves the use of an accurate device model for simulation of small and large signal device and amplifier characteristics. The use of macro-models, or packaged parameter sets based on internal, primarily physics-based equations, is a popular choice especially with regards to CMOS power amplifier design. Following all of the device measurements, characterization and test-structure de-embedding, the preliminary device modeling method involved the use of a BSIM3-based macro-model. In this approach, the model extraction and optimization was performed across temperatures ranging from -25 to 80 degrees Celsius and for device sizes varying from 20 to 60 μm. The extrinsic parasitic elements were added in the core model as scalable functions of temperature and device width. The model performance is verified through measurements of single transistors at various temperatures and for various device sizes. This is an extension of the basic BSIM3-core model extracted for millimeter-wave applications in [19].

The first step in this modeling approach is to measure the DC Ids-Vds curves and S-Parameters (0-65 GHz) across a temperature range (-25 to 80 degrees Celsius) for each of three device sizes of interest (20, 40 and 60x1 μm). For these temperature (T) and device width (W) combinations, a modified BSIM3 model developed for room temperature operation was optimized in Agilent ADS to achieve a fit the measured DC and small-signal measurements, across the entire frequency range. A circuit schematic is shown in Figure 2.8.

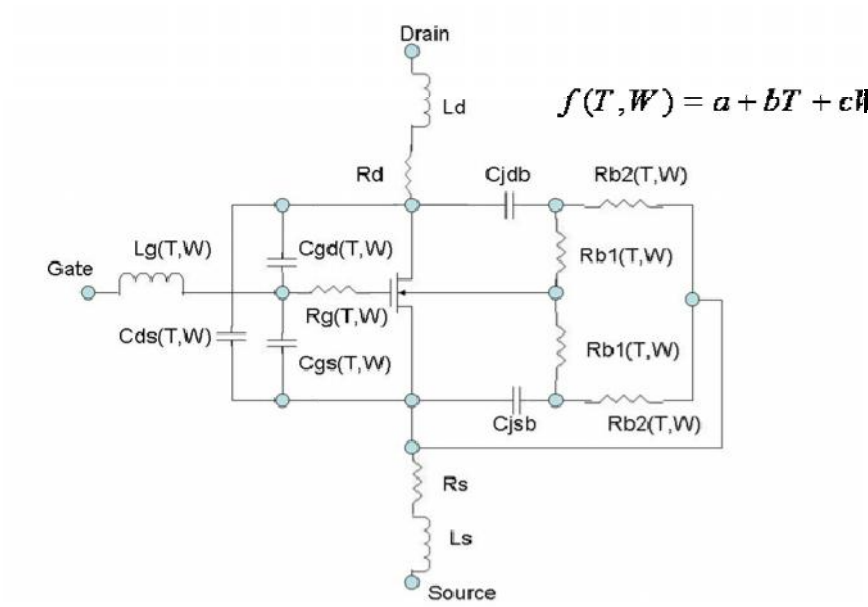


Figure 2.8. Schematic of modified BSIM3-based circuit model.

In Figure 2.8, the BSIM3-based core model is represented by the transistor symbol, and it is comprised of parameter set used to fit the measured DC I-V characteristics. The program IC-CAP provided by Agilent, is used to extract and optimize the BSIM3 parameter set based on device measurements. Furthermore, S-Parameter measurements are used by the same program to optimize the external parasitic values. The optimizing routines are usually not very accurate, so extensive manual optimization is necessary with this approach. Extrinsic parasitics (*Cds*, *Cgd*, *Cgs, Rg, Rb,* etc.) and key internal BSIM3 model parameters such as the Channel Drain- Source Capacitance (*cdsc*) are plotted as functions of T and W. Figures 2.9, 2.10 and 2.11 show examples of these functions under bias conditions Vds=0.9V and Vgs=0.6V with either T or W held constant.

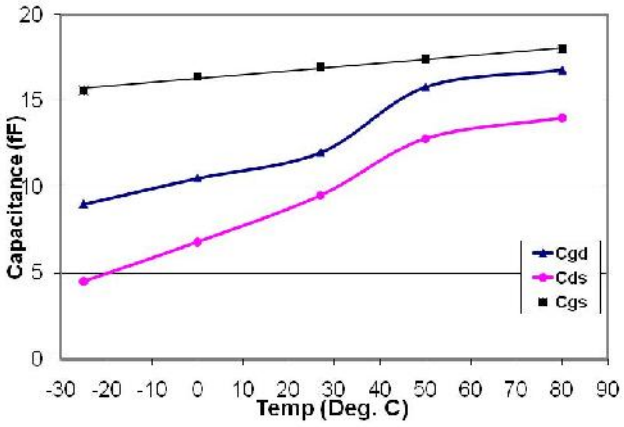


Figure 2.9. Device Capacitances versus Temperature.

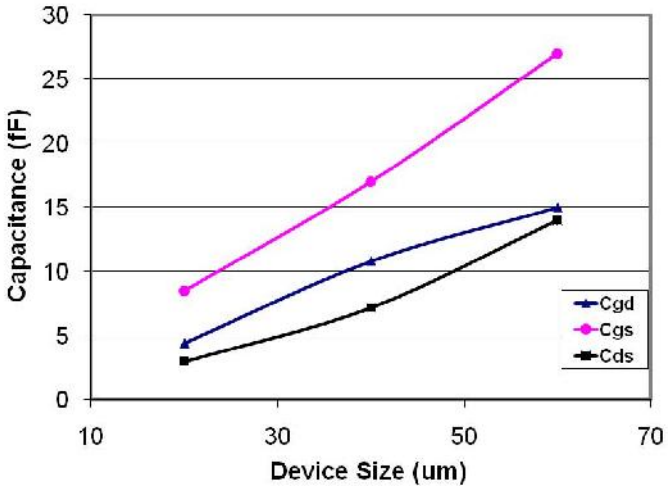


Figure 2.10. Device Capacitances versus Device Size.

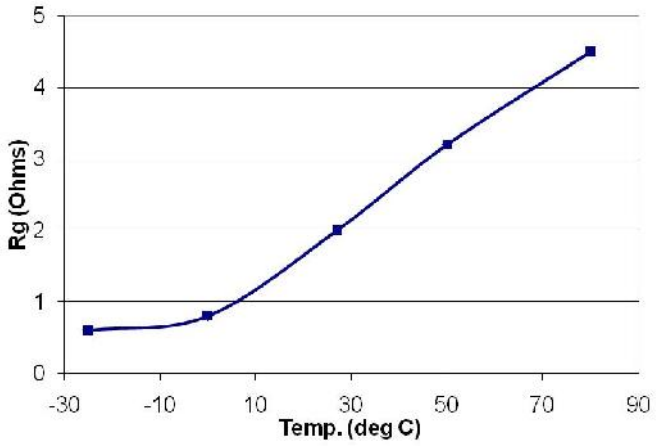


Figure 2.11. Gate resistance versus temperature.

Two-variable second-order polynomial functions for each of the parasitics and model parameters were formed as follows, and implemented in the model :

$$F(T,W) = a + bT + cW + dT^2 + eW^2 + fTW \qquad (2.5)$$

These model functions, in which a, b, c, d, e and f form a unique set of coefficients for each parasitic element and the critical internal parameters, are implemented in the circuit shown in Figure 2.8. With these modifications, the BSIM3 model is modified from what has been previously shown in [19] and the device models have their substrate shorted to the source region.

In Figure 2.12, the temperature variation of the magnitude of S21 at 60 GHz with width held constant at 40 μm is shown, with excellent fit between model and measurements.

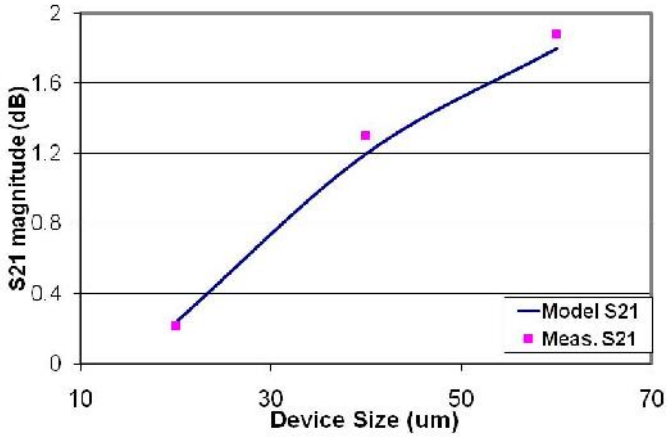


Figure 2.12. S21 magnitude versus device size (Model simulation vs. measured).

In Figure 2.13, the S21 magnitude variation with device size at 60GHz, with temperature held constant at 27 degrees Celsius, is shown, again with good model agreement. In all of the plots shown, Vds=0.9V and Vgs=0.6V. These are both examples of good prediction of the small-signal characteristics using this approach.

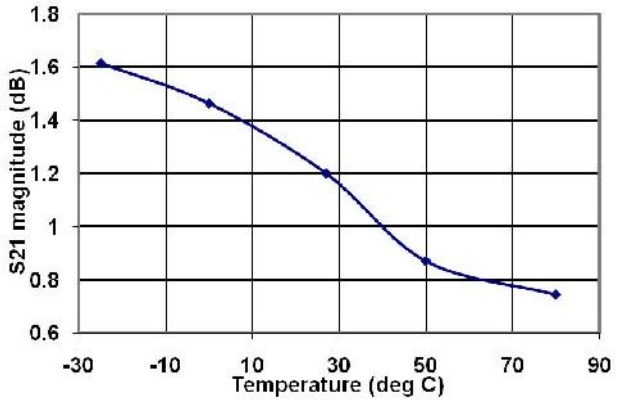


Figure 2.13. S21 magnitude versus temperature (Model simulation vs. measured).

One of the limitations of this approach was found when using the BSIM3-based core model, not in small or large signal discrete transistor simulations, but rather in multi-stage power amplifiers, in which each stage employed the BSIM3-based model. Specifically, convergence issues were often encountered when using the Agilent ADS harmonic balance simulator to simulate these three and four stage power amplifiers. Harmonic balance convergence issues can be observed in any number of unwanted simulator outputs. Examples of these range from non-continuous or non-smooth large signal power sweeps as shown in Figure 2.14 to abortion of the simulation prior to any results being shown. In some cases, the small signal gain of the power amplifier would be simulated accurately, however there would be a discontinuity between this and the predicted saturated output power, Psat. It is important in power amplifier design to have a model which is robust in large signal simulations, otherwise in first-pass design attempts, it is highly difficult to ascertain whether the simulation should be ignored or is indeed an

artifact of the design. Such a guessing game can hardly be afforded in the high-cost tape-out process.

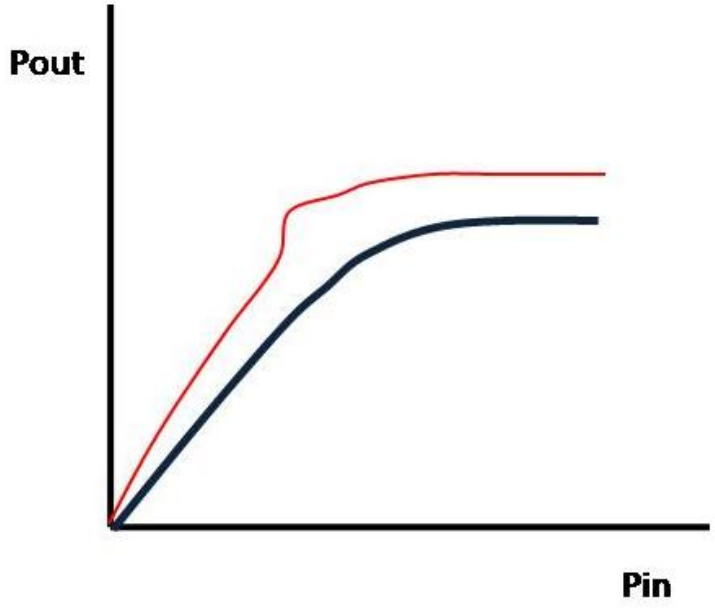


Figure 2.14. Example of a harmonic balance convergence problem encountered with the BSIM3-macromodel (top curve) vs. the expected power amplifier large signal power characteristic.

Other limitations with this modeling approach include problems with its use in load-pull simulations. Load-pull simulations are critical when determining the power, efficiency and gain target impedances that need to be presented to the output of the transistor and when these simulations are not accurate, yet another guessing game ensues. One of the reasons why this model does not lend itself to load pull simulations is the inaccurate simulation of the maximum drain-to-source current of the FET, or Imax. Without a realistic upward limit on this value, as was found in DC I-V simulations, accurate loadline values cannot be calculated. These values are tied very closely with the load pull simulations of the optimum load impedances.

This method, while using a packaged model with built-in parameters, still requires extensive enhancement, as shown, to produce a model suitable for RF amplifier development. The intrinsic and extrinsic parasitic elements still need to be extracted and optimized and placed around the BSIM3-based macromodel core. So given the limitations of this approach which have been encountered, the next step was to seek alternative large-signal modeling methods for CMOS devices to be used in millimeter-wave applications. As will be discussed at length in the next chapter, the focus turned to developing an empirical-based compact model that can be used to accurately predict the millimeter-wave performance of these same devices in both the small and large signal regimes. In essence, the BSIM3 packaged core model is replaced by a novel current generator based solution, the goal being to have a more robust and reliable device model for multi-stage CMOS power amplifier design targeted for millimeter-wave operation. Inspiration for embarking on this approach comes from several models and methods for many different families of semiconductor devices, ranging from Gallium Arsenide MESFETs to HBTs. Models and methods such as those developed by Curtice and Angelov in [20] and [21] are examples of the works consulted during the development of an empirical large signal model suitable for millimeter-wave CMOS power amplifier development.

CHAPTER 3

TEMPERATURE-DEPENDENT CMOS LARGE SIGNAL MODEL FOR MILLIMETER-WAVE APPLICATIONS

3.1 Basic Model Development

The importance of accurate transistor models for millimeter-wave CMOS circuit design has already been well established. The increased focus on millimeter wave CMOS power amplifier development necessitates accurate device models for these applications. When using the industry-standard BSIM-based approach, parasitic elements still need to be added to accurately model mm-wave performance, even following the extensive core model parameter extraction process. Furthermore, as will be demonstrated, the model presented in this work does not exhibit harmonic balance convergence issues when simulating a multi-stage power amplifier under high input drive well beyond the P1dB compression point. This led to the motivation to implement a custom current generator-based model for the short-channel CMOS devices used in mm-wave applications. The parasitic network incorporating size scalability and temperature dependency are also novel features.

A high-level overview of the model development process is shown in the flowchart in Figure 3.1. As shown, static DC and S-Parameter measurements (over the range DC to 65 GHz, if concerned with 60 GHz power amplifier design) are the critical measurements needed in this approach. Furthermore, having a temperature controller as part of the on-wafer measurement setup is optional but necessary for the incorporation of temperature dependency. The nominal sized device is a 40μm periphery device. Device size scalability is added as necessary.

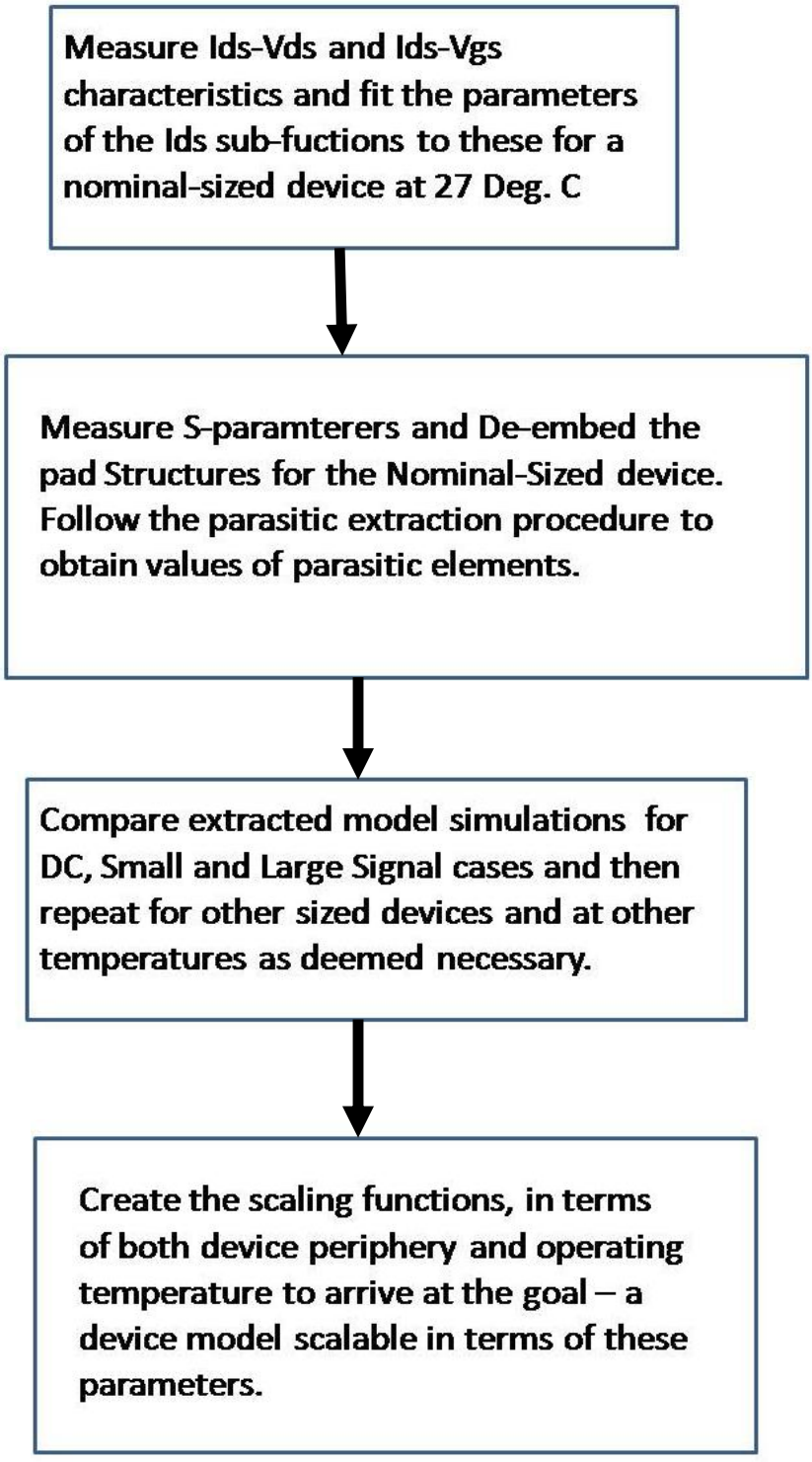


Figure 3.1. Flowchart showing model extraction procedure.

The final circuit schematic of the extracted large-signal model, which was implemented in the microwave circuit simulator, Agilent Advanced Design System (ADS), is shown in Figure 3.2. The realization of the Ids current source and parasitics as functions of the applied voltages and process and temperature parameters will be discussed in the subsequent sections.

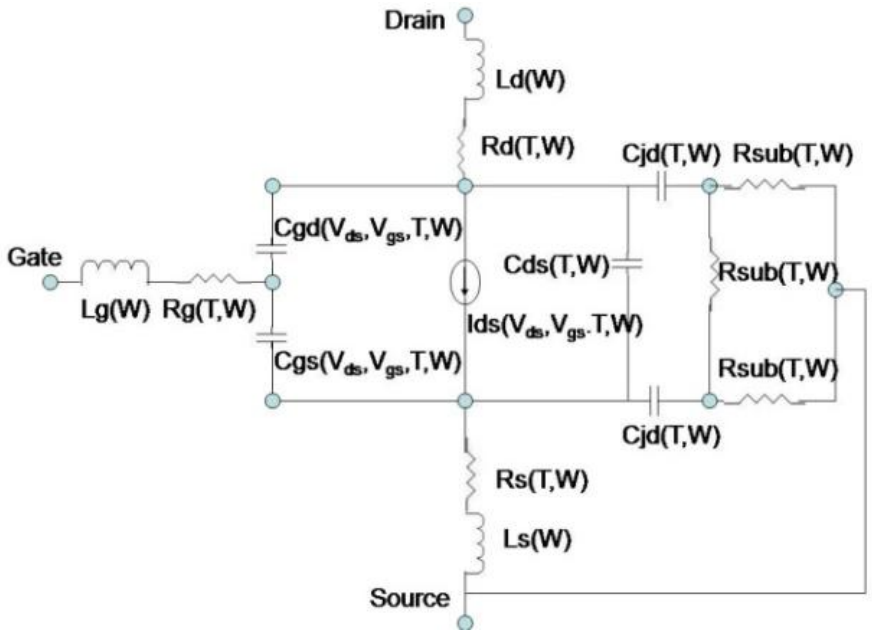


Fig 3.2. Schematic of circuit model with bulk tied to source.

For a simplistic MOSFET which does not take into account short gate length (short channel) behavior, the current source shown in the schematic in Figure 3.2, would be based on equations (3.1) and (3.2) and would model the linear and saturation regions of the I-V plane according to long-channel behavior.

$$I_{DS} = K \cdot \left((V_{GS} - V_T) - V_{DS}/2\right) \cdot V_{DS} \quad \text{for } V_{DS} \leq V_{GS} - V_T \quad (3.1)$$

$$I_{DS} = K \cdot (V_{GS} - V_T)^2/2 \quad \text{for } V_{DS} > V_{GS} - V_T \quad (3.2)$$

$$K = \frac{1}{2}\mu_n C_{ox}\left(W/L\right) \quad (3.3)$$

The boundary between linear and saturation regions at $V_{DS} = V_{GS} - V_T$ represents the onset of channel pinch-off. As V_{DS} increases, the localized voltage between the gate and a point near the drain end of the channel becomes smaller and smaller for a given applied gate voltage. At some point this difference becomes equal to the threshold voltage, V_T, and the onset of saturation occurs for that particular V_{DS}.

However, it is difficult to generate a large-signal model based on these equations because such a model needs to have smooth and continuous transitions between the regions of operation in the I-V plane. Furthermore, for short-channel devices used in millimeter-wave applications, as will be shown later, equation (3.2) is no longer valid and equation (3.1) might not accurately fit the Ids-Vds characteristic in the knee-voltage region . Before, going into the model development, a summary of short-channel CMOS device physics, taken into account during model development, will be presented.

3.2 CMOS Short Channel Device Physics

In this modeling approach, mention has been given to the equations accurately modeling the so-called short channel effects especially present in CMOS field-effect transistors with short gate lengths below 100nm. In this work, devices are used with gate lengths of 90 and 130nm. The device physics relating to these short-channel effects is

presented next. Much of the device physics groundwork for this modeling approach is compiled in [22].

3.2.1 Velocity Saturation

In short-channel FETs, the carriers reach velocity saturation at lower Vds values than in long-channel devices. The general principle of velocity saturation is illustrated in Figure 3.3. Within the device there will be an electric field value for which the velocity will saturate, marked Vsat. This also corresponds to a saturation mobility value.

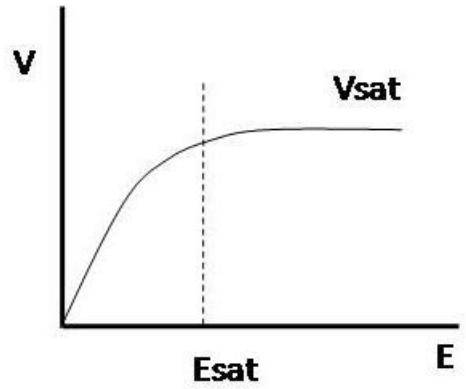


Figure 3.3. Drift Velocity versus lateral Electric field.

The general form of the carrier velocity – electric field relation is given as follows:

$$V = \frac{\mu\varepsilon}{1 + \frac{\varepsilon}{\varepsilon_{sat}}} \qquad (3.6)$$

Note that for ε >> ε_{sat} , the above relation reduces to,

$$V = \mu\varepsilon_{sat} \qquad (3.7)$$

Furthermore, when the channel length decreases into the short-channel regime, the drain-source current in saturation can be written, as in (3.8), in terms of the long channel saturated drain-source current. The latter obeys the well-know "square-law" relation, specifically that Ids is proportional to the square of the quantity, Vgs – Vt.

$$I_{ds,sat} = \frac{long - channel\ I_{ds,sat}}{1 + \frac{V_{gs} - V_T}{\varepsilon_{sat} L}} \qquad (3.8)$$

It is important to see that in (3.8), when $\varepsilon_{sat}L \ll V_{gs} - V_T$, as is the case with short-channel devices, when the square-law numerator is divided by the denominator, the equation will be proportional to $V_{gs} - V_T$ instead of $(V_{gs} - V_T)^2$. This is the fundamental effect that velocity saturation has in short-channel FETs – there is no longer a square-law Vgs dependence in saturation.

3.2.2 Channel Length Modulation

In the case of long-channel FETs, the onset of drain current saturation occurs with pinch-off and Ids no longer varies with increasing Vds. Near the drain side of the channel, the channel is no longer touching the drain and the characteristic triangular

shape (with thinner width) near the drain side is obtained. As Vds keeps increasing, the electric field moves the channel depletion region further and further away from the drain, resulting in incremental decrease in channel length with increasing Vds. This is channel length modulation and this effect is obviously much more pronounced in short-channel FETs. This results in there being a finite conductance in the saturation region of short-channel devices due to the fact that Ids will now increase with Vds as opposed to staying constant in the long-channel case.

This characteristic is observed in the measured and modeled Ids vs. Vds curves previously shown in Figure 3.5.

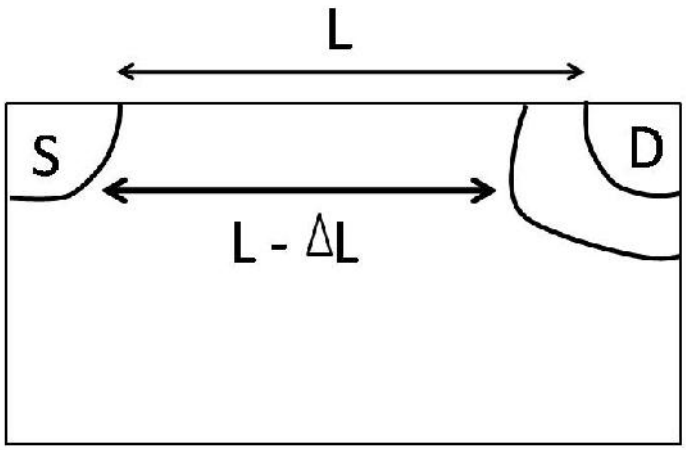


Figure 3.4. Channel length reduction due to increasing Vds in short-channel device.

3.2.3 Drain-Induced Barrier Lowering

This is another short-channel effect, which the model equations that will be introduced, will take into account. The mechanism behind the lowering of the threshold voltage is the lowering of the source-channel and drain-channel barriers with increasing Vds. This

causes a reduction in the threshold voltage with increasing Vds. This can be taken as a linear decrease with respect to Vds, resulting in drain-induced barrier lowering (DIBL) as follows:

$$\mathrm{Vth}^{*} = \mathrm{Vth} - \mathrm{n} \cdot \mathrm{Vds} \qquad (3.9)$$

3.3 Nonlinear Drain Current Source

The nonlinear drain current generator is implemented as in (3.10), utilizing two sub-functions, F_1 and F_2, of Vds and Vgs respectively, shown in (3.11) and (3.12).

$$\mathrm{Ids}(\mathrm{Vgs}, \mathrm{Vds}) = \mathrm{F}_1(\mathrm{Vds}) \cdot \mathrm{F}_2(\mathrm{Vgs}) \qquad (3.10)$$

The natural logarithm as a function of a polynomial of Vds shown in (3.11) is the chosen form because it readily and accurately models the high linear region resistance observed in these short-channel devices, in addition to providing a smooth transition to the saturation region.

$$\mathrm{F}_1(\mathrm{Vds}) = \ln\left[1 + \left(\frac{1}{2} + \frac{1}{2}\tanh(\mathrm{b}_1\mathrm{Vds})\right) \cdot (\mathrm{a}_1\mathrm{Vds} + \mathrm{a}_2\mathrm{Vds}^2 + \mathrm{a}_3\mathrm{Vds}^3)\right] \qquad (3.11)$$

$$\mathrm{F}_2(\mathrm{Vgs}) = \left(\frac{1}{2} + \frac{1}{2}\tanh\left(\mathrm{b}_2(\mathrm{Vgs} - \mathrm{Vth}^{*})\right)\right) \cdot \left(\mathrm{b}_3 + \mathrm{b}_3 \cdot \tanh\left(\mathrm{b}_4(\mathrm{Vgs} - \mathrm{b}_5)\right)\right) \cdot \exp\left[-\ln\left(1 + \exp\left(\frac{-\mathrm{Vgs} + \mathrm{Vth}^{*}}{\mathrm{nV_T}}\right)\right)\right] \qquad (3.12)$$

$$\mathrm{Vth}^{*} = \mathrm{Vth} - \mathrm{n} \cdot \mathrm{Vds} \qquad (3.13)$$

The sub-functions F_1 and F_2 model the Ids-Vds and Ids-Vgs characteristics shown in Figures 3.5 and 3.6 respectively.

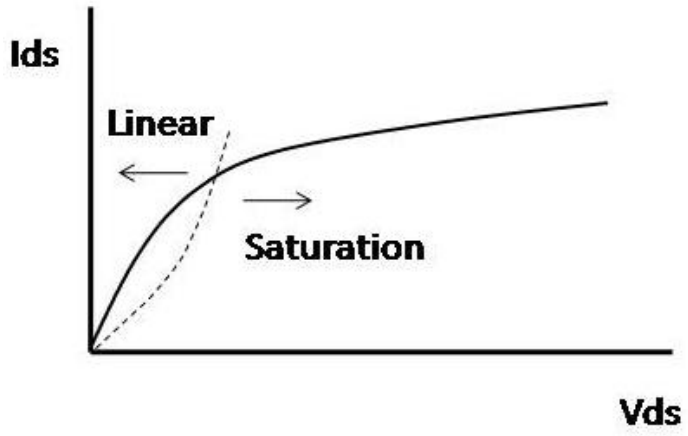


Figure 3.5. Representative Ids vs. Vds curve for a fixed Vgs value in strong inversion.

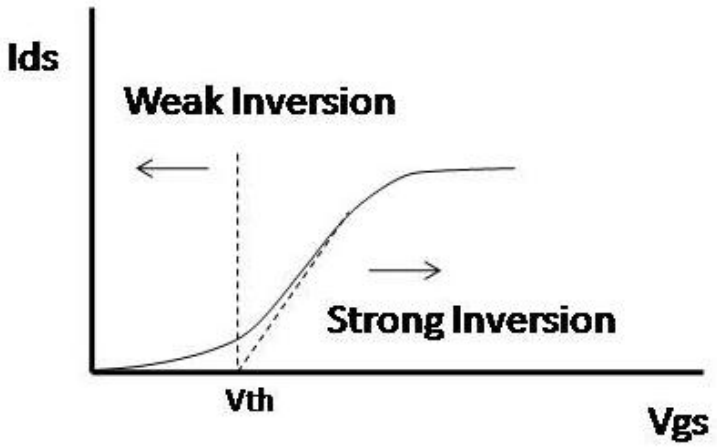


Figure 3.6. Representative Ids vs. Vgs curve for a fixed Vds value.

Table 3.1 Description of parameters in Ids model equations.

PARAMETER	PURPOSE/FUNCTION
a^1, a^2, a^3	Polynomial coefficients for Ids-Vds curves modeling linear and saturation regions continuously.
$b1$	Slope parameter in Heaviside Step Function used to eliminate discontinuity at Vds=0.
$b2$	Slope parameter in Heaviside Step Function used between strong and weak inversion in Ids-Vgs equation.
$b3, b4, b5$	Coefficients/Shifting variables for the hyperbolic tangent function used to model Ids-Vgs in strong inversion region

In both (3.11) and (3.12), hyperbolic tangent-based Heaviside step functions, such as the first term in (3.12), are used for smooth transitions between regions of the I-V plane. The general form of the Heaviside step function is as follows and will be explained below:

$$H(x) = \frac{1}{2} + \frac{1}{2} \cdot \tanh\left(b \cdot (x - x_1)\right) \qquad (3.14)$$

In all of the model equations, whenever a smooth, continuous transition is needed, the Heaviside step function is used to achieve this. It is used in both of the Ids sub-functions as well as when incorporating temperature-dependency of the current source as will be

subsequently shown. The standard hyperbolic tangent function varies from +1 to -1, as an odd function about zero. To achieve a smooth, continuous step function, the Heaviside approach scales this function by ½ and then shifts it up by ½, thereby creating a step function between zero and one on the y-axis. The variable, x_1, in (3.14) is used to shift the transition of the step function anywhere along on the x-axis. For illustration purposes, the function is plotted in Figure 3.7 for various values of the parameter "b". This parameter shown in equation (3.14) can be increased or decreased to create a more abrupt or more gradual transition between regions. In all cases, about an arbitrary x-value, for increasing x, the function will tend to unity and for decreasing x, will tend to zero, which is the desired outcome for a step function. The staircase step shown in Figure 3.7 is attainable as the variable b tends to infinity.

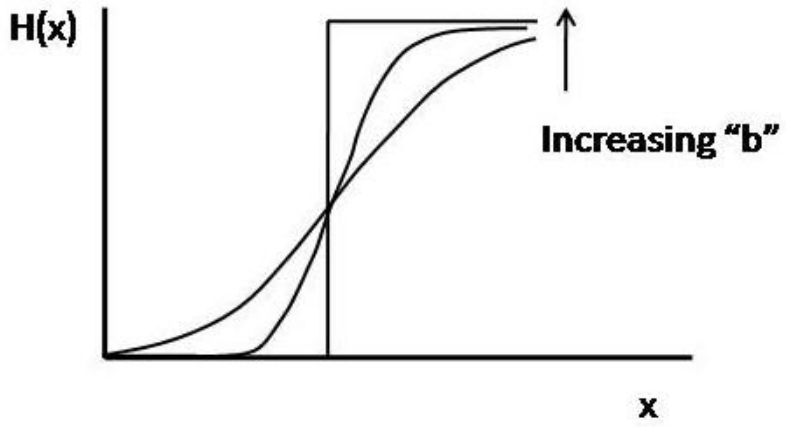


Figure 3.7. Heaviside Step Function for various parameter values.

To demonstrate the accuracy of this model, the simulated and measured DC I-V characteristics are plotted together for an 80x1 μm device in Figure 3.8.

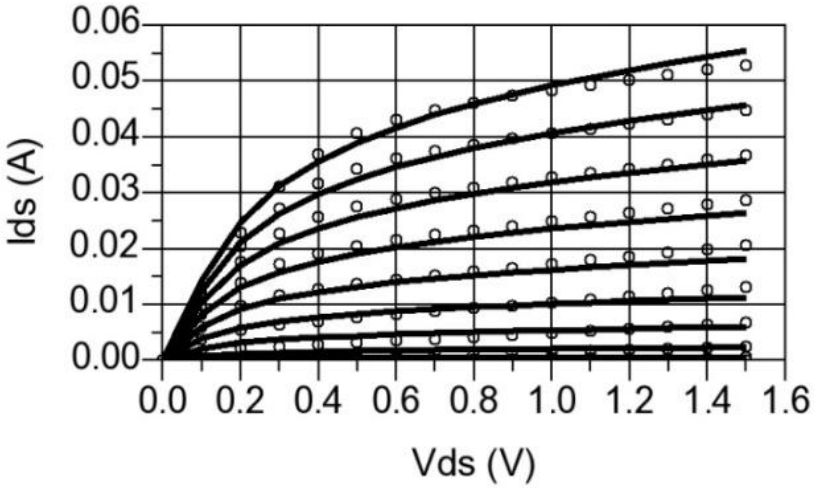


Figure 3.8. Ids vs. Vds (0-1.5V) for Vgs (0 to 1.0V) for an 80x1 μm, 90nm NMOS device.

Excellent agreement between measured and simulated Ids versus Vgs and transconductance (gm) versus Vgs for the same periphery device are shown in Figures 3.9 and 3.10.

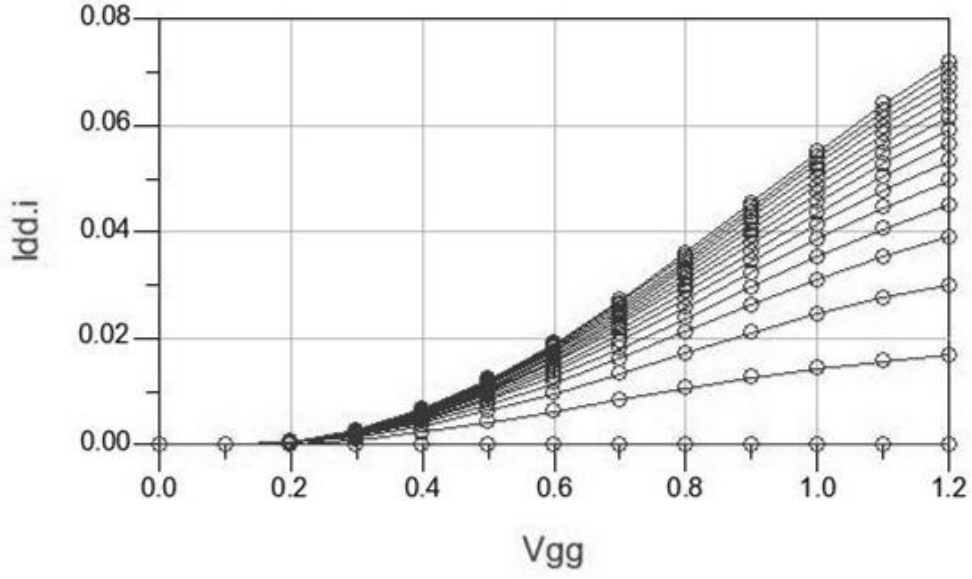


Figure 3.9. Ids vs. Vgs (0-1.5V) for Vgs (0 to 1.0V) for an 80x1 μm, 90nm NMOS device.

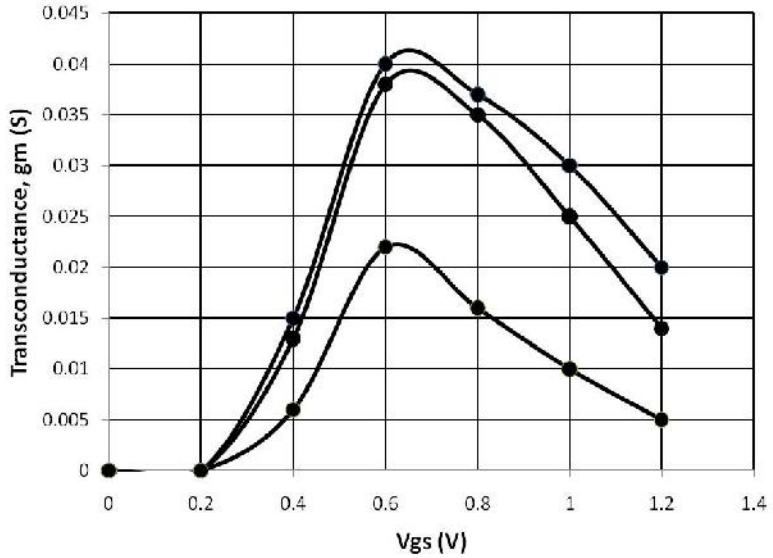


Figure 3.10. gm, vs. Vgs (Vds=0.6, 0.9 and 1.2V) for 80µm device.

Now an in depth description of the sub-functions will be given. Referring back to the F_1 sub-function in (3.11), for sub-100nm gate-length transistors, the measured Ids has a slow rising slope with respect to Vds in the area of the knee voltage, a well-known short channel CMOS characteristic. A novel natural logarithm-based expression as a function of a cubic polynomial, having a shallow slope and continuous gradual curvature around the knee voltage is used to accurately model Ids-Vds. The hyperbolic tangent as function of Vds tends to zero close to Vds=0 and the natural logarithm function subsequently becomes zero because the unity term dominates for any Vds values less than this. For any values of Vds greater than zero, the polynomial expression as a function of Vds dominates and with proper selection of the coefficients a1, a2 and a3 the natural logarithm becomes a function of this polynomial exclusively and the slow rising nature of the natural logarithm function allows for accurate modeling of the short-channel Ids-Vds characteristic in the knee region. The polynomial function within the natural logarithm function also has the use of providing the Ids-Vds slope in the saturation region. This function is also continuously differentiable in Vds. The hyperbolic tangent-based term in

(3.12) has coefficients chosen such that as Vgs keeps increasing in the saturation region, there is a Vgs – Vth dependence that is closely approximated and which is observed for short channel devices.

In the Vgs sub-function, F_2, shown in (3.12), the exponential-natural logarithm term incorporates the subthreshold Ids-Vgs characteristic with continuous transition to the strong inversion Ids-Vgs characteristic, modeled by the second term in the product. Vth* as given in (3.13) is used to implement the Drain-Induced Barrier Lowering effect. In equation (3.12), the part of the subthreshold equation for MOSFETs dependent on Vgs is given, in which W and L are the width and length of the transistor respectively, q is the coulomb charge constant, T is the temperature. This part of the equation is directly from the work presented in [23]. It is necessary to have a physical subthreshold expression incorporated into the device model because many CMOS applications such as subthreshold CMOS receivers [24] rely on subthreshold model accuracy. Another important component that relies on accuracy below Vgs cutoff is the Doherty amplifier which, in the most conventional implementation, uses a peaking amplifier biased in Class C mode.

The incorporation of the DIBL parameter into the Vth term in equation (3.13) leads to the Ids increasing with increasing drain voltage for a given gate voltage in the subthreshold region. A plot of the measured and simulated Ids versus Vds for Vgs values of 0.1 and 0.2V is shown in Figure 3.11.

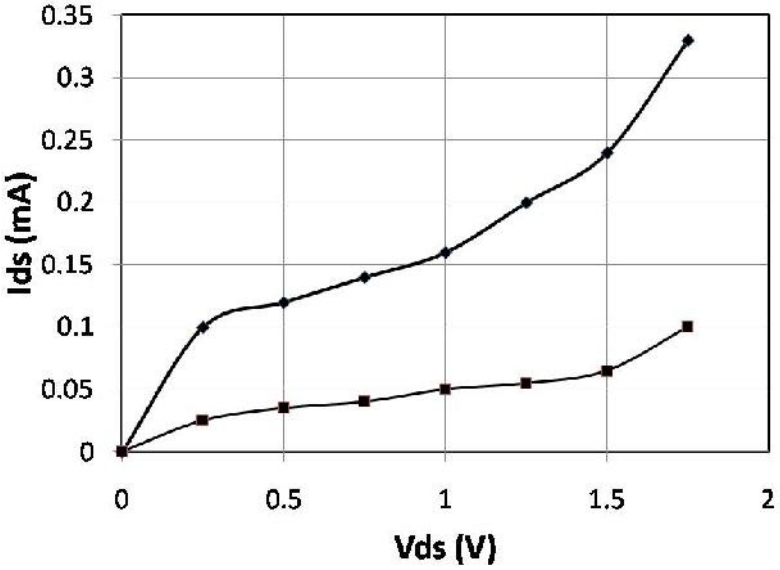


Figure 3.11. Ids vs. Vds for Vgs=0.1 and 0.2V for a 80μm device.

When the current generator is optimized fully, it is implemented as a symbolically defined device (SDD) in Agilent ADS. The method of implementing an SDD-based model is similar to that found in [25]. A screenshot of the circuit schematic from ADS is shown in Figure 3.12. Note that the box in the center connected to the gate, drain and source nodes (left, top and bottom nodes respectively), is the SDD and is a function of two applied voltages, Vgs and Vds. Parasitic element scalability is not included in this schematic to simplify the view. A listing of the drain current equation parameters with their optimized numerical values is provided in Table 3.2. These are values for a 40μm device operating at room temperature. These are the nominal values of device periphery and operating temperature, as will be discussed in section 3.5. However, for ease of use of this large signal model in a circuit simulator such as ADS, the default values for the current generator will be those listed in Table 3.2. The user can optimize these to fit the characteristics of a similarly-sized device in the process of choice.

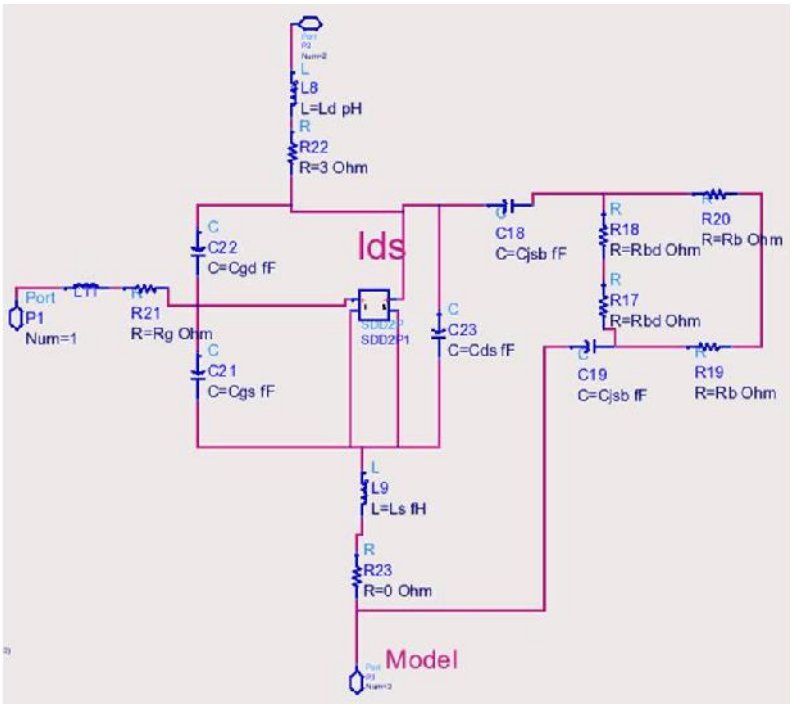


Figure 3.12. Agilent ADS implementation of basic version of circuit model with the SDD- based drain current generator.

Table 3.2 Numerical values of drain current equation for 40µm, 90nm CMOS device.

Parameter(s)	Optimized Numeric Value(s)
a_1, a_2, a_3	50, -2, 3
b_1	10
b_2	2.5
b_3, b_4, b_5	0.4, 0.4, 1.8

3.4 Circuit Model Extraction for Millimeter-wave Applications

This current generator is implemented as a symbolically-defined device in Agilent ADS along with the extracted parasitics to form the complete circuit model shown in Figure 3.2. One fundamental difference between this approach and others such as those presented in [20] and [26] is that the parasitics are implemented as scalable functions of temperature, T, and total device width, W, as will be described later. First an overview will be given of the process by which the parasitics are extracted from measured and de-embedded device S-Parameters.

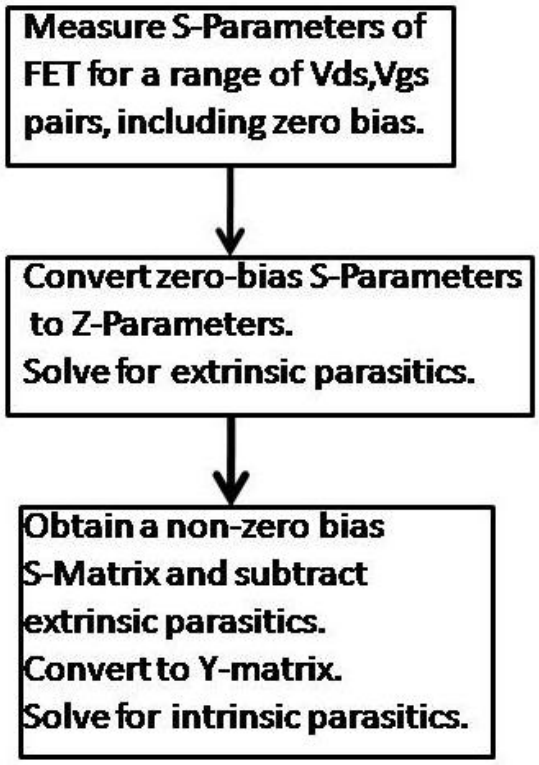


Figure 3.13. Model Extraction Flowchart

As shown in the flowchart, the FET is biased at complete zero bias (Vgs=0 and Vds=0), thereby shutting off the transconductance current source and leaving an effective network of the series extrinsic parasitics. These S-parameters are converted to Z-parameters and they are written as shown in (3.14) – (3.16). The method for extraction of the extrinsic parasitic is similar to that presented in [27].

$$Z_{11} = R_s + R_g + \frac{R_c}{3} + j\omega(L_s + L_g) \qquad (3.14)$$

$$Z_{12} = R_s + \frac{R_c}{2} + j\omega L_s \qquad (3.15)$$

$$Z_{22} = R_s + R + R_c + j\omega(L_s + L) \qquad (3.16)$$

Note that in these equations, Rc, is the channel resistance and is a process-dependent parameter, obtained from the process manual. With this known, it is readily observed that Rs is the difference of the real part of Z_{12} and Rc. Ls is given by the imaginary part of Z_{12}. Then Rg is obtained from the real part of Z_{11} in (3.14) and Lg from the imaginary part of the same equation. In a similar fashion, Rd and Ld can be obtained from the real and imaginary parts of Z_{22} in equation (3.16). Now that the extrinsic parasitic are known, they can be used to obtain the Y-parameters of the intrinsic device. The device S-Parameters are measured at a chosen bias point (Vgs, Vds) and then are converted to Z-Parameters. The series parasitics obtained are then subtracted from the Z-Parameters using (3.14) to (3.16) as a guide. The resulting Z-matrix is then converted to a Y-matrix. In this approach, the extracted gate resistance, Rg, is the only series extrinsic element not subtracted, resulting in it being found in the Y-parameters of the intrinsic device as observed in (3.17) through (3.20).

$$Y_{11} = \frac{\omega^2(C_{gs}+C_{gd})^2 R_g + j\omega(C_{gs}+C_{gd})}{1+\omega^2(C_{gs}+C_{gd})^2 R_g} \quad (3.17)$$

$$Y_{12} = \frac{-\omega^2 C_{gd}(C_{gs}+C_{gd})^2 R_g - j\omega C_{gd}}{1+\omega^2(C_{gs}+C_{gd})^2 R_g} \quad (3.18)$$

$$Y_{21} = \frac{g_m}{1+j\omega(C_{gs}+C_{gd})R_g} \quad (3.19)$$

$$Y_{22} = \frac{j\omega C_{jd}}{1+j\omega C_{jd} R_{sub}} + j\omega(C_{ds}+C_{gd}) + \frac{\omega^2 C_{gd}^2 R_g + j\omega g_m C_{gd} R_g}{1+j\omega(C_{gs}+C_{gd})R_g} \quad (3.20)$$

Previous model development approaches such as [28] derive the Y-Parameters as above and that is why they are left in this form. However, one of the key differences in extraction for mm-wave applications occurs when solving for the intrinsic parasitics (capacitances and substrate resistances). Specifically, lower-frequency approximations (valid below 10 GHz) are made, such as:

$$\omega^2(C_{gs} + C_{gd})^2 R_g \ll 1 \quad (3.21)$$

To optimize for model performance through 65 GHz, the equations (3.14)-(3.20) are solved without such approximations, yielding:

$$C_{gd} = -_{\text{I}}\mathrm{Re}(Y_{11})/(\mathrm{Im}(Y_{11}) \cdot \omega \cdot R_g)] \cdot [\mathrm{Im}(Y_{12})/\mathrm{Im}(Y_{11})] \quad (3.22)$$

$$C_{gs} = \mathrm{Re}(Y_{11})/(\mathrm{Im}(Y_{11}) \cdot \omega \cdot R_g)] \cdot [1 + \frac{\mathrm{Im}(Y_{12})}{\mathrm{Im}(Y_{11})}] \quad (3.23)$$

Following the extraction of these parasitics, the drain-source capacitance and substrate resistance and capacitances are extracted using the expressions derived in (3.22) and (3.23) and are given below:

$$C_{jd} = \frac{Im(Y_{22}) + Im(Y_{12})}{\omega} \tag{3.24}$$

$$C_{ds} = \frac{Im(Y_{22})}{\omega} - C_{gd} - C_{jd} - g_m R_g C_{gd} - R_{sub} C_{jd} \tag{3.25}$$

$$R_{sub} = Re(Y_{22})/(\omega^2 {C_{jd}}^2) \tag{3.26}$$

Furthermore, for the capacitances C_{gd} and C_{gs}, continuous functions are formed to fit a sufficient number of capacitance values extracted at discrete bias points similar to the method presented in [29]. Plots of these capacitance-voltage functions are shown in Figures 3.14 and 3.15 for a 40µm device.

\

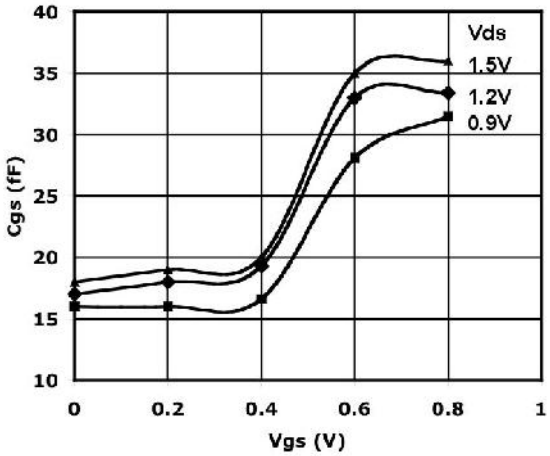


Figure 3.14. Cgs vs. Vgs for varying Vds for a 40µm device.

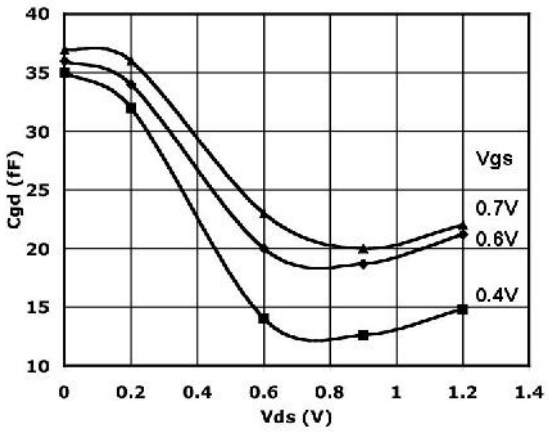


Figure 3.15. Cgd vs. Vds for Vgs = 0.4 and 0.6V for a 40μm device.

The equations used to fit these extracted C-V curves are shown in (3.27) and (3.28). Note the use of the hyperbolic tangent functions in fitting these characteristics.

$$C_{gs} = \{C_{gs0} + [e_o + e_1 \tanh(V_{gs} - e_2 \cdot V_{th})][e_3 + e_4 \tanh(V_{gs} - V_{ds}]\} \quad (3.27)$$

$$C_{gd} = C_{gd0} + [f_o + f_1 \tanh(V_{gs} - V_{ds}) - f_2 V_{th})] \quad (3.28)$$

Furthermore, the capacitances Cgs and Cgd are implemented as voltage-variable functions utilizing the symbolically-defined devices in ADS. The procedure for properly doing this is not as simple as inserting nonlinear capacitors into the model in place of static value capacitors. Since the goal is to use symbolically-defined devices (SDDs) in ADS to implement these capacitances, it is necessary to generate current-based equations, since SDDs are typically implemented as current sources as functions of voltages applied to them. For example, the nonlinear drain current source is implemented as an SDD which is a function of Vgs and Vds.

When generating current equations it is generally not appropriate to simply take the time derivative of the product of the voltage-varying capacitance directly, as this can introduce large errors in modeling the voltage variation of the capacitance in the model [30]. Rather, the capacitance-voltage expressions, however formulated, should be integrated to produce charge-voltage expressions which can then be differentiated with respect to time. The SDDs in ADS allow for differentiation with respect to the time variable by simply selecting a model parameter. So all that remains is to insert a charge-voltage expression into the SDD and select the appropriate SDD parameter index. In general form, the capacitances Cgs and Cgd are expressed in this approach as functions of the applied voltages as shown in (3.27) and (3.28). Qgs can be obtained by integrating Cgs with respect Vgs and Qgd can be obtained by integrating Cgd with respect to Vgd (or Vgs – Vds). Since both capacitance expressions are based on hyperbolic tangent based functions, integration of these nonlinear capacitance expressions with respect to Vgs or Vgd, whichever applies, will yield charge equations, Qgs and Qgd of the following general form:

$$Q(v) = \int_0^v \left(c_0 + c_1 \tanh(v - v_0)\right) dv$$
$$= c_0 v + c_1 v_1 \ln\left(\cosh(v - v_0)\right) \qquad (3.29)$$

Creating SDDs using such functions and then differentiating with respect to time is non-problematic because these sorts of functions are continuous and are handled by the nonlinear simulator [30].

The device capacitances, in addition to the resistances and inductances are extracted for a range of device peripheries. Plots are shown for some of these circuit model elements as functions of device size in Figures 3.16, 3.17 and 3.18.

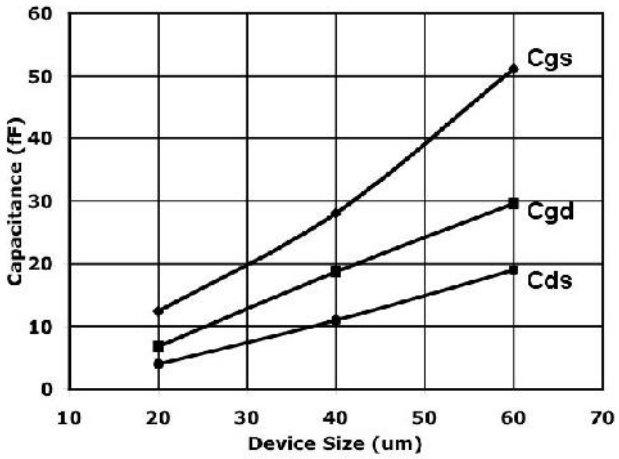


Figure 3.16. Model capacitances versus device size.

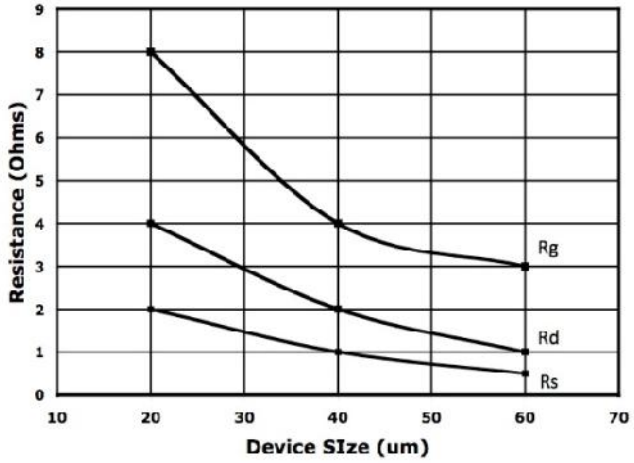


Figure 3.17. Model Rg, Rd and Rs as functions of devices size.

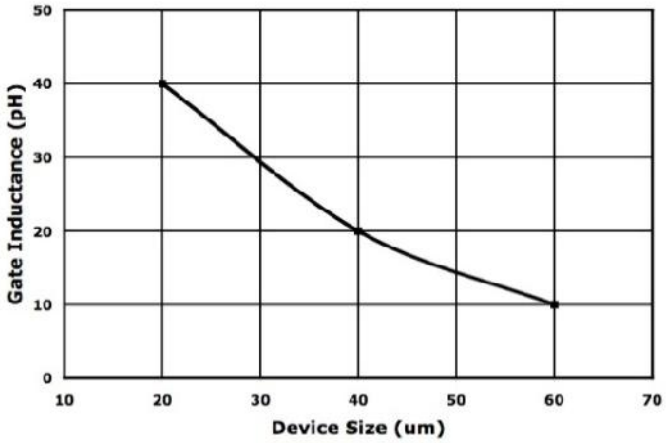


Figure 3.18. Model gate inductance as function of device size.

The resistances and inductances all have an inverse relationship with increasing device size or periphery. This is expected because the total device size is proportional to the number of gate fingers (as well as source and drain connections) in parallel. Adding more of the fingers in parallel leads to an overall decrease in the total resistances and inductances. Conversely, the total capacitance scales fairly linearly as an increasing function of total device width.

Additionally, some of the parasitic elements, especially the resistances, exhibit temperature variation. An example is the variation of the extracted gate resistance for a 40μm device, shown in Figure 3.19. The model has been extracted for various device sizes and at various temperatures, however the next step is to create a unified model that scales in terms of device periphery and that is temperature dependent as well. The approach to achieving this is described in the next sub-section.

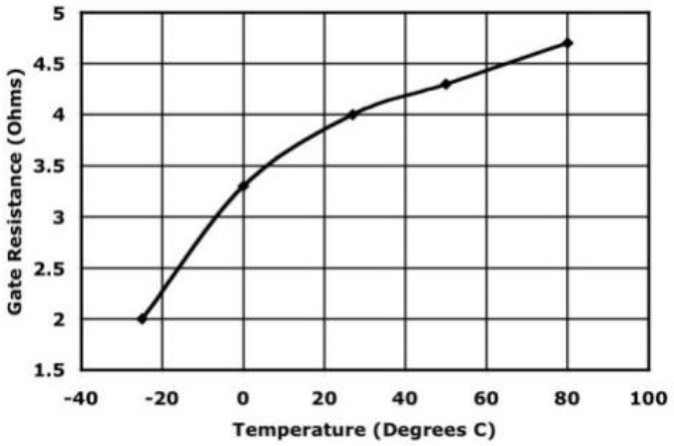


Figure 3.19. Gate resistance vs. temperature for a 40μm device.

3.5 Implementation of Temperature-Dependency and Size Scalability

Since these devices will be used in amplifiers integrated on chips which will be used in variable temperature environments, the approach presented incorporates temperature scalability by first measuring the drain current of devices over a wide range of temperatures (-25 to 80 degrees Celsius) and establishing a temperature-scalable function. A temperature-dependent drain-current plot is shown in Figure 3.20. An alternate plot in Figure 3.21 better illustrates the inversion in the drain current variation with temperature with respect to the applied gate-source voltage. Some of the variations of the CMOS device characteristics with process and temperature scaling are given in [31].

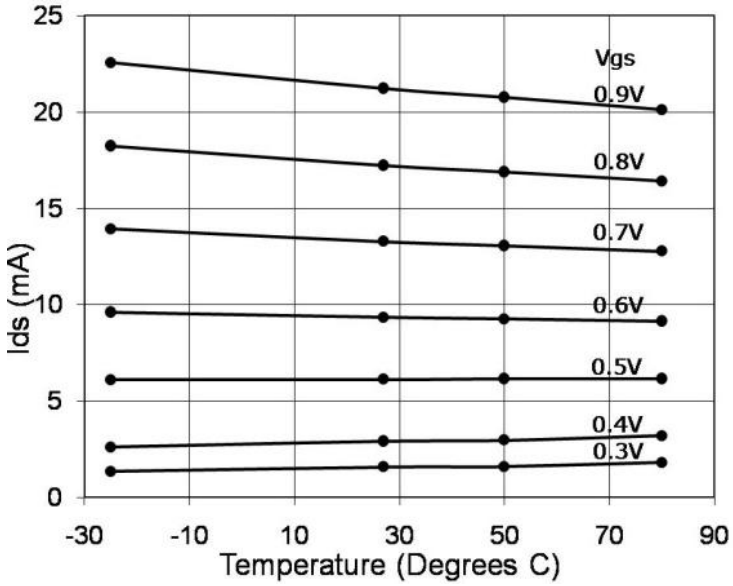


Figure 3.20. Ids vs. temperature and Vgs for Vds= 1.0V for a 40x1 µm, 90nm NMOS device.

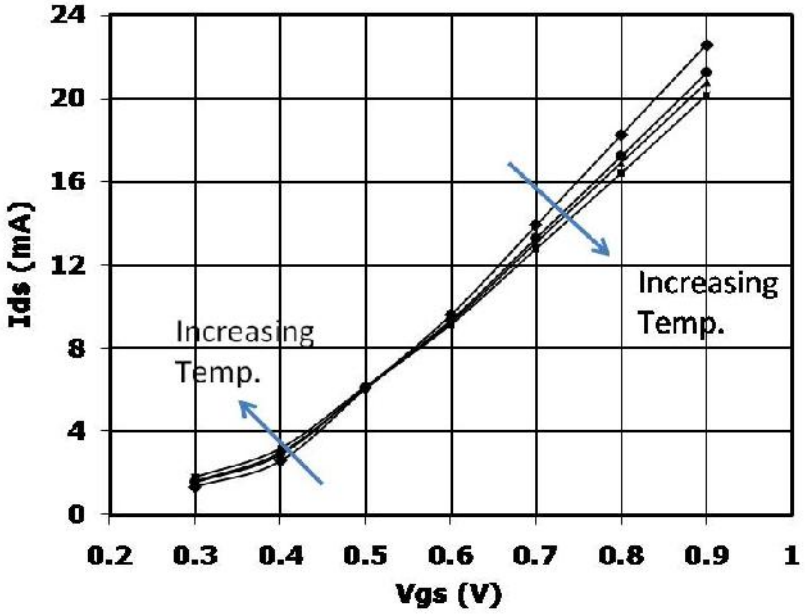


Figure 3.21. Ids vs. Vgs for various temperatures (0 to 80 degrees Celsius) for a 40x1µm, 90nm NMOS device.

A scalable function in terms of temperature and Vgs is established, taking the form:

$$f(V_{gs}, T) = 1 + \beta \cdot \left[\tanh\left(\alpha \cdot (V_{gs} - \gamma \cdot 0.5)\right)\right] \cdot (t_1 \cdot T + t_2 \cdot T^2 + t_3 \cdot T^3) \qquad (3.30)$$

In (3.30), the temperature variation is modeled by a third-order polynomial and the hyperbolic tangent term is used to model the inversion at Vgs=0.5V with respect to how Ids varies with temperature (decreasing versus increasing slope). The basis for this function is observed in Figures 3.20 and 3.21. For a given applied drain and gate voltage, a general third order polynomial can be formed as a function of temperature to model the temperature variation of the drain current. However, as an example, for a 40um device with Vds=1.2V, there is an inversion which is observed at Vgs=0.5V. Specifically, for Vgs values greater than 0.5V, there is an inverse relationship with temperature and there is a direct relationship between the drain current and temperature for Vgs less than 0.5V. The hyperbolic tangent function centered at Vgs=0.5V is offset in magnitude between positive and negative values of β and is used to model this inversion with respect to the applied Vgs. The magnitude of the hyperbolic tangent function is set by β, and the steepness about the transition voltage, Vgs=0.5V for a 40μm device, is controlled by the parameter α. Note the similarity to the Heaviside step function used in the nonlinear drain current generator. In addition, the parameter γ is formed as a function of Vds. It should be noted that the polynomial function of temperature will scale with device width when multiplied by the device-size scalable drain current function.

An important point to understand is that equation (3.30) is multiplied by the general Ids equation optimized at the nominal temperature, 27 degrees Celsius, to produce a temperature-dependent function across a desired temperature range. This works because

the third-order polynomial has coefficients chosen such that it is equal to zero at 27 degrees Celsius, becoming increasingly negative for higher temperatures and vice versa. A plot of this function is shown in Figure 3.22. When this function is multiplied by the hyperbolic tangent function and then added to unity, it results in a reduction in drain current for temperatures greater than nominal and vice versa for Vgs values greater than 0.5V. Similarly it results in an increase in drain current for temperatures greater than the nominal temperature and vice versa for smaller Vgs values. The polynomial function allows for accurate modeling of greater negative slope of Ids versus temperature as illustrated in Figure 3.22 for greater values of Vgs.

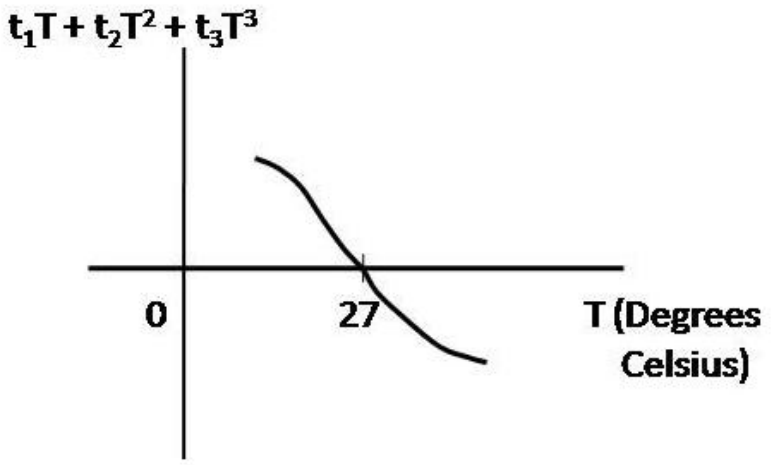


Figure 3.22. Third-order polynomial as a function of temperature used in temperature-dependent implementation of Ids.

The observed inversion can be explained by the lessening mobility reduction as a function of increasing temperature observed for lower values of Vgs. Furthermore, for values of Vgs less than about 0.5V, the increase of Ids with greater temperature can be attributed to the slight lowering of the threshold voltage, Vth, with increasing

temperature. This reduction of Vth has much less of an effect at higher values of Vgs, such as Vgs=1.0V, and the scattering mechanisms leading to mobility reduction have a greater effect with rising temperature. A summary of the parameters used in the drain current temperature-scaling function is presented in Table 3.3.

Table 3.3 Summary of parameters used for incorporation of temperature dependency into drain current equation.

Parameter(s)	Purpose/Function
$\frac{r(s}{\alpha}$	Parameter used to adjust the slope of the transition between positive and negative values of the tanh function used to model inversion.
β	Scaling parameter used to optimize the tanh function used to model inversion.
γ	Scaling parameter to adjust center Vgs value.
t_1, t_2 and t_3	Coefficients of third-order polynomial used to control drain current variation with temperature.

The general scattering mechanisms leading to changes in the mobility with respect to the electric field applied to the device are illustrated in Figure 3.23.

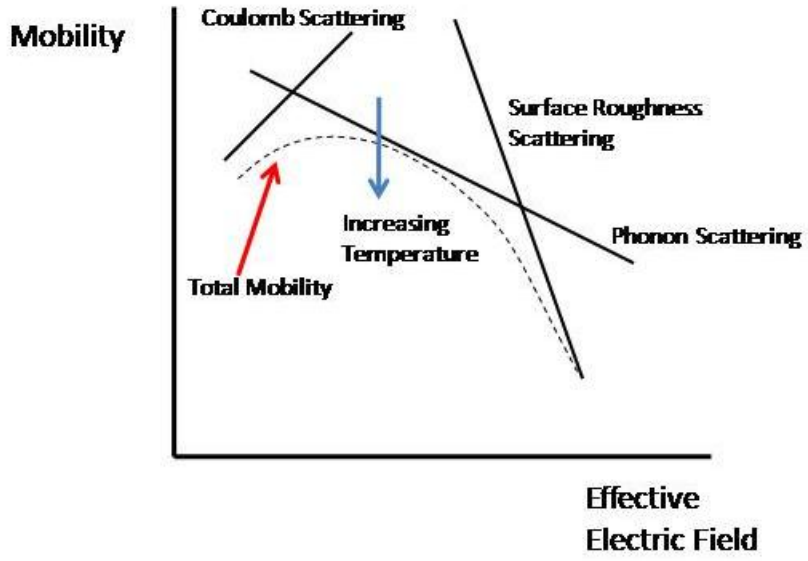


Figure 3.23. Mobility vs. electric field characteristic in a FET with scattering mechanisms which form the characteristic.

In Figure 3.24, it is shown that with this temperature dependency in addition to device width scaling added to the drain current equation as in (3.31), the model accurately simulates the S21 magnitude variation with temperature.

$$\mathrm{Ids}(T, W, \mathrm{Vgs}, \mathrm{Vds}) = f(V_{gs}, T) \cdot \alpha \cdot \left(\frac{W}{W_0}\right) \cdot (\mathrm{Ids}_0) \quad (3.31)$$

In (3.32)-(3.35), Ids_0, R_0, C_0 and L_0 are the nominal values extracted at T=27 degrees Celsius and W=40 µm. It was observed that the resistances and capacitances change most

with temperature and are similarly implemented as temperature-dependent functions as in (3.32) and (3.33).

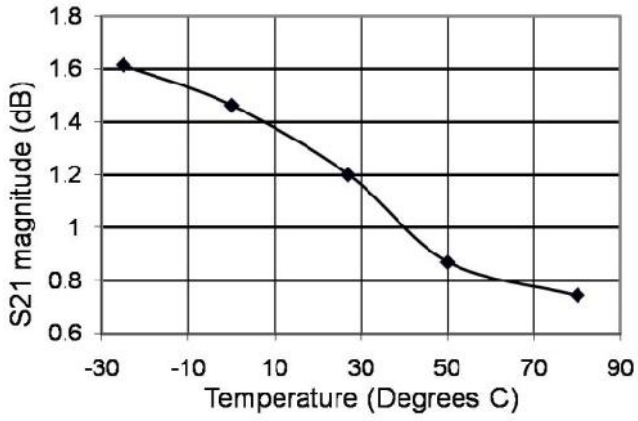


Figure 3.24. Simulated vs. measured S21 vs. temperature for 40x1 µm device for Vds=1.0V, Vgs=0.6V.

As an improvement to the method shown in [20], in addition to the drain current scaling, the critical parasitics are extracted across a range of temperatures and device peripheries (20, 40, 60 and 80 µm). To more accurately model parasitic variation across the entire temperature range, a polynomial scaling function is used as opposed to the linear approach shown in [26].

$$R(T,W) = \beta \cdot \left(\frac{W_o}{W}\right) \cdot (r_1 T + r_2 T^2 + r_3 T^3) \cdot R_o \quad (3.32)$$

$$C(T,W) = \gamma \cdot \left(\frac{W}{W_0}\right) \cdot (c_1 T + c_2 T^2 + c_3 T^3) \cdot C_0 \quad (3.33)$$

$$L(W) = \delta \cdot \left(\frac{W_0}{W}\right) \cdot L_0 \quad (3.34)$$

To demonstrate size scalability, for 40x1 and 80x1 μm transistors, simulated S21, S11 and S22 are compared with measurements across the frequency range DC to 65 GHz.

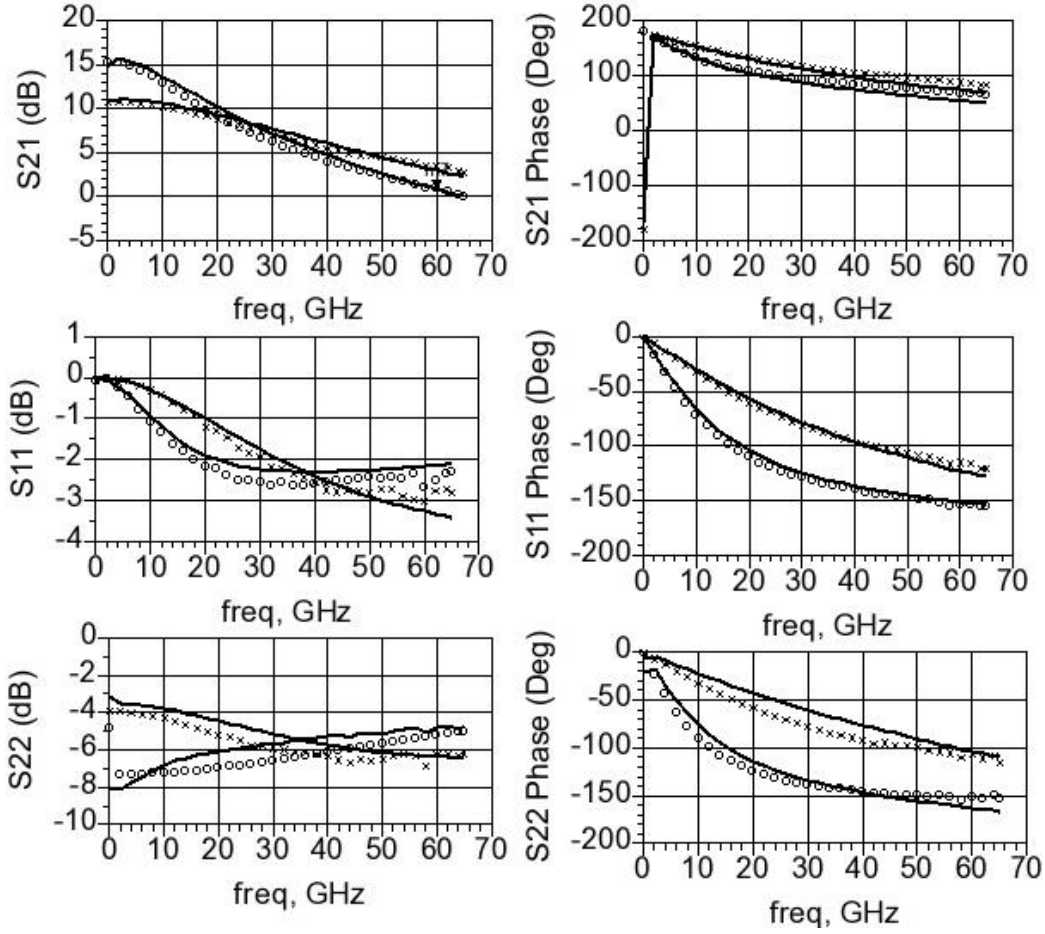


Figure 3.25. Simulated (solid) vs. measured magnitude and phase of S21, S11 and S22 (circles for 80x1um and x-marks for 40x1um device) at fixed temperature, 27C.

In Figure 3.26, simulated and measured large signal power performance is shown at the device level for 40 and 80μm periphery devices. In addition to verifying device small-signal performance, a large signal power sweep was conducted for both 40 and 80μm devices at 60GHz at room temperature and compared with measurements. Temperature-dependent large signal power performance is demonstrated in the next

chapter when the temperature-dependent model is used in a four-stage power amplifier and performance is shown for different temperatures.

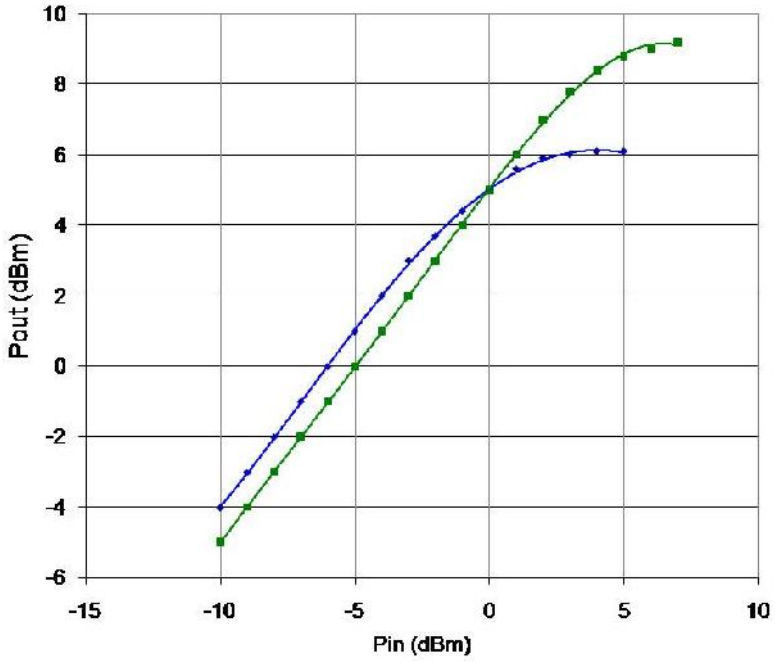


Figure 3.26. Measured and simulated large signal power sweeps for 40µm (bottom blue curve) and 80µm (top green curve) devices at room temperature, 27 degrees Celsius.

CHAPTER 4

60GHZ CMOS POWER AMPLIFIER DESIGN

With the CMOS device model validated for small and large signal performance through 65GHz for a range of device peripheries and across a broad range of temperatures, it is ready for use in multi-stage 60 GHz CMOS power amplifier design. Such designs, using single-ended architectures, are presented in [32] and [33]. The first part of the design process involves choosing suitably sized device periphery for the required output power as well as the number of stages needed for achieving the required power gain. Then the optimum load and source impedances for the each of these transistor sizes are determined. This forms a starting point for developing the matching networks (input, output and interstage) for the power amplifier. There are a couple of approaches that can be used to arrive at the optimum load and source impedances. The first method involves performing load and source pull simulations on the large signal model to arrive at the optimum impedances for maximum output power, power-added efficiency (PAE) and gain. In this work, Agilent ADS software was used for this purpose and an example test bench is shown in Figure 4.1. Note that this is a generic test bench from an ADS example project and does not reflect the frequencies of interest nor the device model used. The bench consists of a load tuner in this example, but can easily be modified to include a source tuner. Only the impedance at the fundamental frequency is varied and the impedances at the harmonic frequencies are all set to 50Ω.

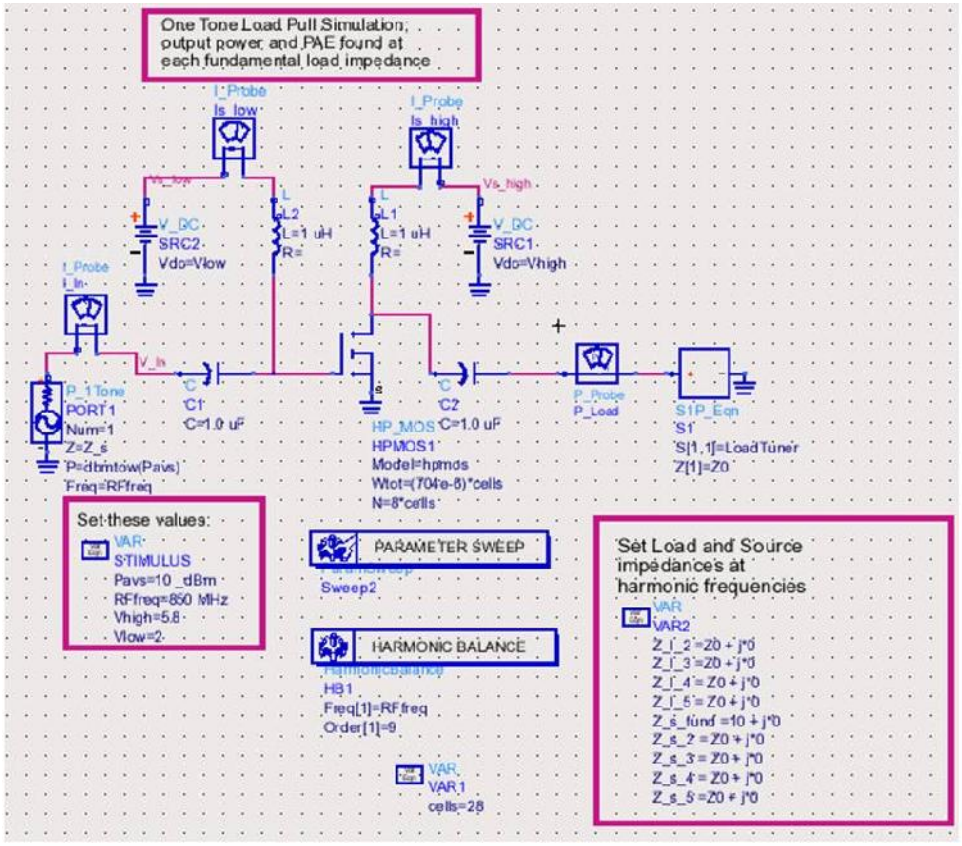


Figure 4.1 Example of Load Pull simulation bench in Agilent ADS.

When using the developed large signal model in a load/source pull simulation bench such as the one shown, the optimum power and efficiency load impedances are shown on the Smith Charts in Figures 4.2 and 4.3 for 40µm and 80µm devices respectively at 60 GHz. These plots correlate with what would be expected – decreasing real part of the optimum load impedance with increasing device periphery.

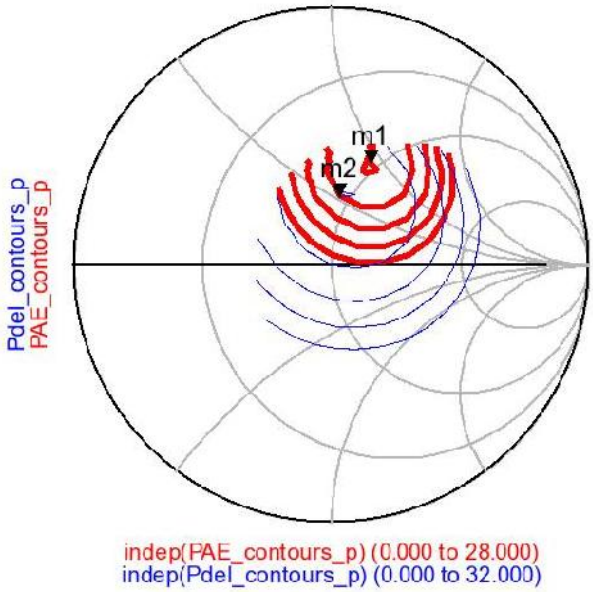


Figure 4.2 Power and PAE contours for a 40µm NMOS transistor at 60 GHz based on developed large signal model. Maximum delivered power = 6.2dBm. Maximum PAE=24%. Vdd=1.0V and Vgg=0.6V.

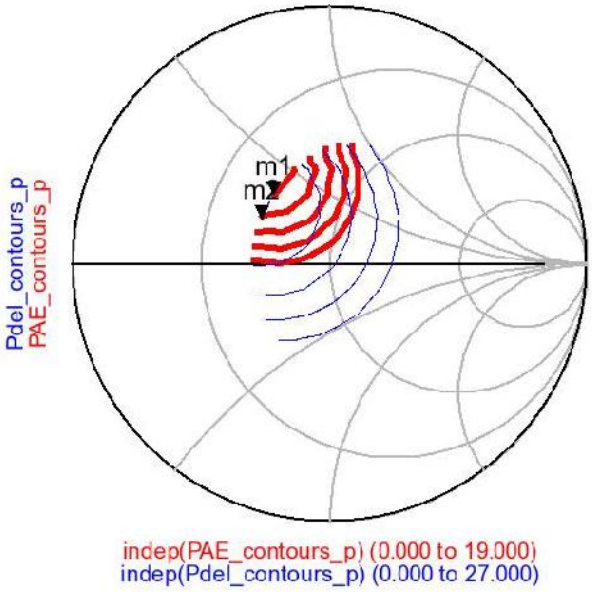


Figure 4.3. Power and PAE contours for an 80µm NMOS transistor at 60 GHz based on large signal model. Maximum delivered power = 9.1 dBm. Maximum PAE=29%. Vdd=1.0V and Vgg=0.6V.

The optimum load and source impedances to be presented to the FET at 60GHz for maximum output power and gain respectively are plotted in the Smith Chart in Figure 4.4 for a range of device peripheries.

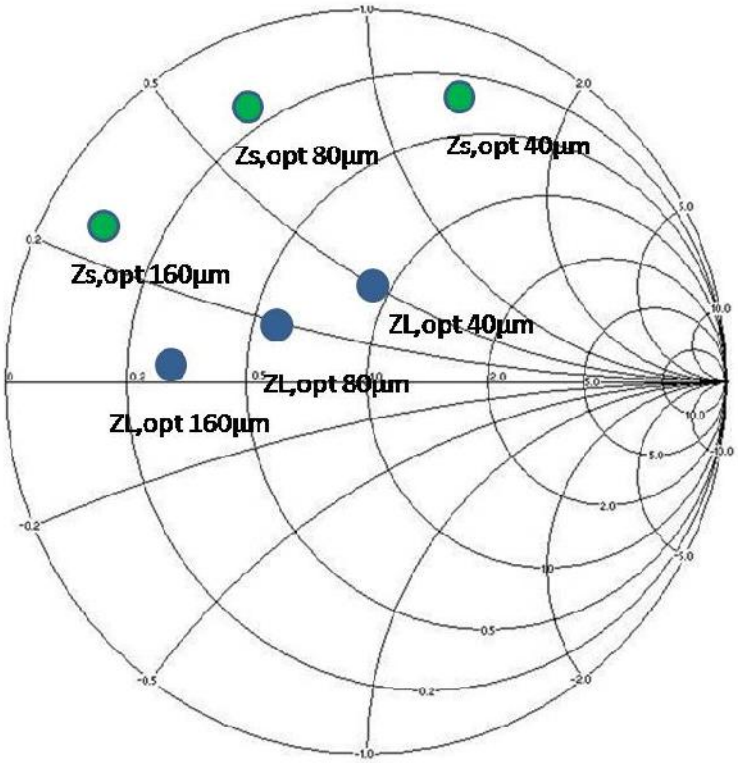


Figure 4.4 Optimum load and source impedances for 40, 80 and 160µm devices at 60GHz.

The other, more fundamental method of determining the optimum load impedance involves use of the device's I-V curves, the corresponding load line for maximum power and the total output capacitance of the device. Since all of the stages of the 60GHz power amplifiers will be biased in Class AB mode (between Class A and Class B), calculating the Class A loadline resistance to determine Ropt is a good starting point. This loadline is shown in Figure 4.5.

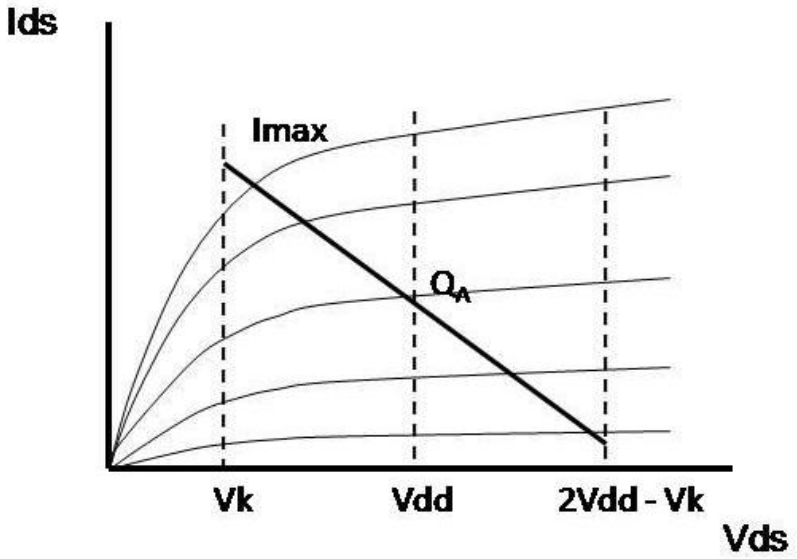


Figure 4.5. DC Loadline for maximum output power for a FET with Class A quiescent bias.

This graph takes into account the device knee voltage which is significant in a short-channel NMOS transistor. Note that the calculated load line resistance value would actually be the same for a Class B power amplifier as it would for the Class A case due to the fact that under both quiescent biases, the amplifier is still capable of producing the same maximum output power just with 6dB more input drive power required in the case of Class B. These fundamentals of power amplifier design are presented in [34] and [35].

The value of the optimum resistance to be presented to the output of the FET can be calculated directly as:

$$R_{opt} = \frac{2 \cdot (V_{dd} - V_k)}{I_{max}} \qquad (4.1)$$

The maximum output power can be expressed as:

$$P = \frac{1}{2} \cdot (V \quad - V_k) \cdot \left(\frac{I_{max}}{2}\right) \qquad (4.2)$$

However, noticing that one can substitute (4.1) into (4.2), the following expression is obtained in terms of only Vdd, Vk and Ropt:

$$P = \frac{(V_{dd} - V_k)^2}{2 \cdot R_{opt}} \qquad (4.3)$$

Now that Ropt is known, the next step is to find Copt, which will have a negative value and a magnitude equal to the device output capacitance. When the Ropt and Copt, in parallel, are presented to the output of the FET, the Copt will cancel the device output capacitance and Ropt will be presented to the output of the FET. The output matching network (or an interstage matching network) will need to perform an impedance transformation to the impedance formed by the parallel combination of Ropt and Copt. Copt can be estimated as the sum of the device capacitances Cds and Cgd. As an example, for a 40μm periphery device, Ropt will be equal to approximately 52Ω and Copt will be equal to -0.046pF (-46fF). This approach is useful because these values are typically normalized and thus can be easily used for matching all of the stages of a multistage power amplifier consisting of various device peripheries. For this set of devices, the normalized Ropt and Copt will be equal to 2.1Ω·mm and -1.16pF/mm respectively.

Using these methods, three and four-stage CMOS power amplifiers were designed using the large-signal model with targeted performance at 60GHz. First a flowchart is presented in Figure 4.6, illustrating the overall power amplifier design flow. The first three steps have been discussed in previous chapters. The design steps are in the highlighted boxes.

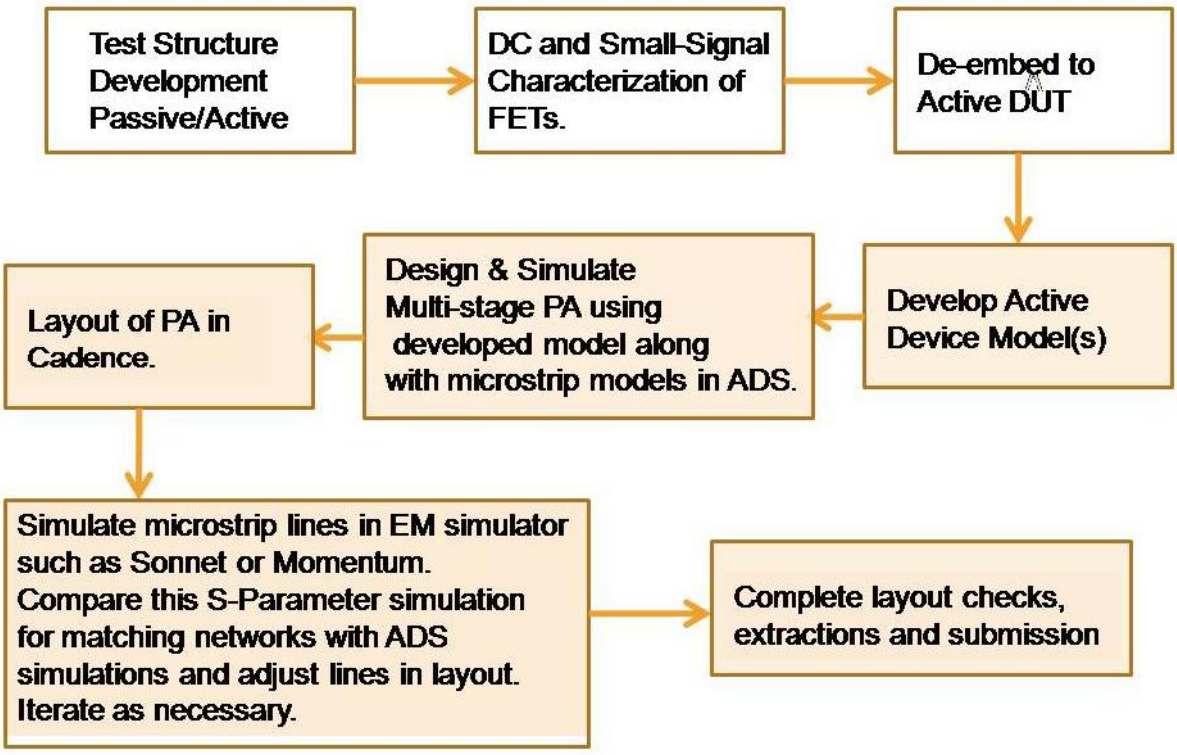


Figure 4.6. Power amplifier design flow.

Both of the designs utilize amplifier stages which are all biased in Class AB mode. The three-stage PA was designed and taped-out first, using the 90nm CMOS ST Microelectronics process. A schematic and die photos are shown in Figure 4.7.

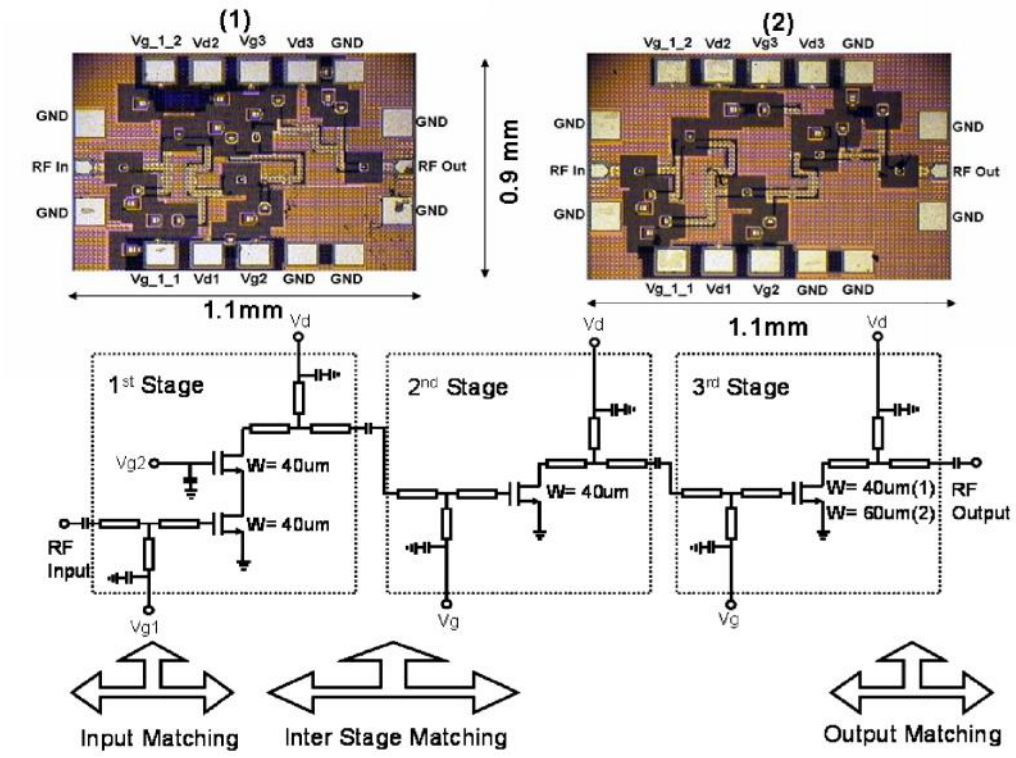


Figure 4.7. Schematic of three-stage 60GHz CMOS power amplifier.

The design approach started with the design of the output power stage to achieve a saturated output power (after the output matching network losses) of around 8dBm when using a 60µm device in the output stage. A second version using a 40µm device was also designed, however the measured saturated output power was around 3dB lower, suggesting less optimal matching for this amplifier in addition to using a smaller device in the output stage. In both cases, two interstage matching networks were implemented. One was between the optimal source impedance of the output stage and the Zopt of the second stage FET. The first interstage match was between the source impedance necessary to match the second stage FET and an optimal gain match for the first stage cascode structure. The cascode structure, consisting of a common-source input device and common-gate device, was used to provide increased isolation of the input device as

well as potentially higher bandwidth around 60GHz, due to reduction of the Miller capacitance. The measured and simulated small and large signal power performance for the three stage PA with 60µm output stage are shown in Figures 4.8 and 4.9.

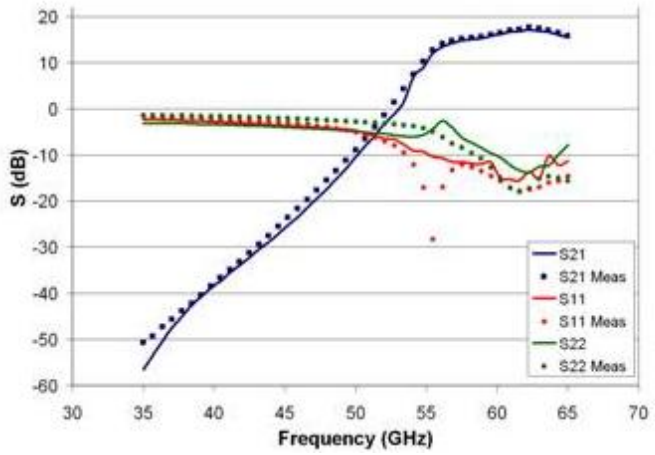


Figure 4.8. Measured and simulated S-Parameters of three-stage 60GHz CMOS PA.

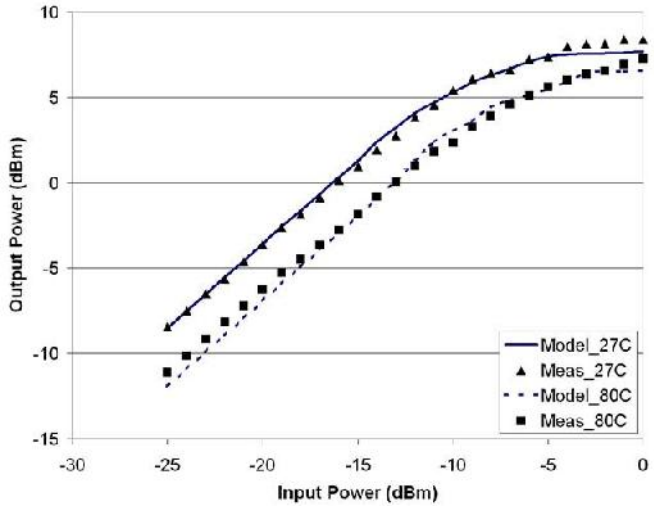
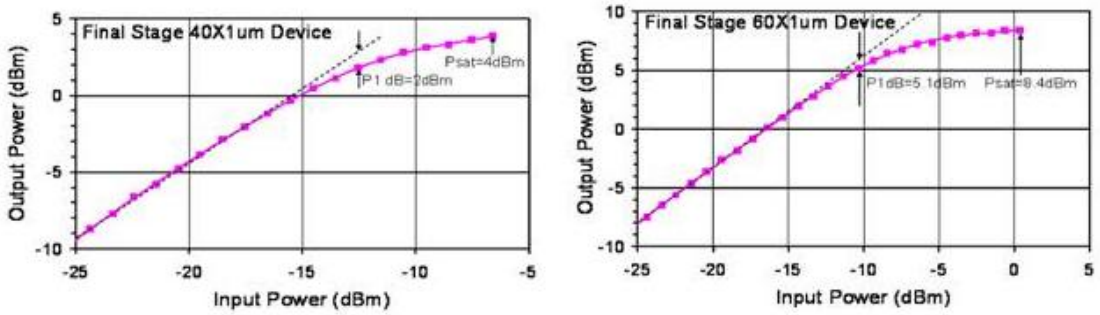


Figure 4.9. Measured and simulated power performance of three-stage 60GHz CMOS PA over temperature.

A comparison of the output power of the two versions of the three-stage PA is given in Figure 4.10 and a performance summary for the 60µm version is presented in Table 4.1.



(a) (b)

Figure 4.10. Comparison between large signal power performance of three stage PA with 40 and 60µm output stages.

Table 4.1. Performance summary of three-stage 60GHz CMOS PA.

Gain	17dB
3dB Bandwidth	57-65GHz
Psat	8.4 dBm
P1dB	5.1 dBm
DC Power Consumption	54 mW
Input/output match	< -12 dB
Chip area	0.9 mm^2

In both the three and four-stage 60GHz designs, the matching networks were implemented with microstrip lines formed by using the top metal layer for the signal layer on top of the bottom M1 layer. Millimeter-wave low-loss microstrip transmission lines were used instead of coplanar lines to achieve a more compact layout. In both the three and four stage designs, series-shunt-series matching networks were used. In both cases, gate and drain bias (usually of around 0.6-0.8V and 1.0-1.2V respectively) were applied directly through DC bias pads with metal-insulator-metal (MIM) bypass capacitors to ground at the point of bias application. The MIM-capacitors to ground were also used following the shunt stubs for an RF path to ground. RF chokes were not used in any of these designs. The schematic of a four-stage CMOS PA is shown in Figure 4.11, the main difference being a 160μm output stage to deliver around 13.5dBm of saturated output power. The extra stage gives an increased gain of 20dB as compared with 17dB for the three-stage designs.

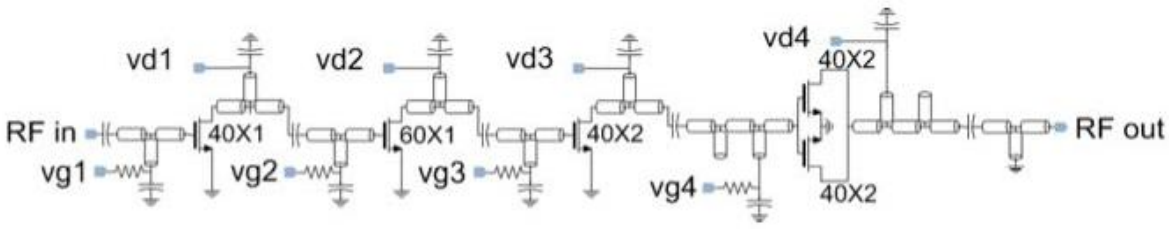


Figure 4.11 Schematic of 4-stage power amplifier

A die photo of the fabricated chip is shown in Figure 4.12 and measured and simulated small and large signal performance are shown in Figures 4.13, 4.14 and 4.15, demonstrating robust model accuracy. The accurate prediction of drain current with input power drive is demonstrated in Figure 4.16.

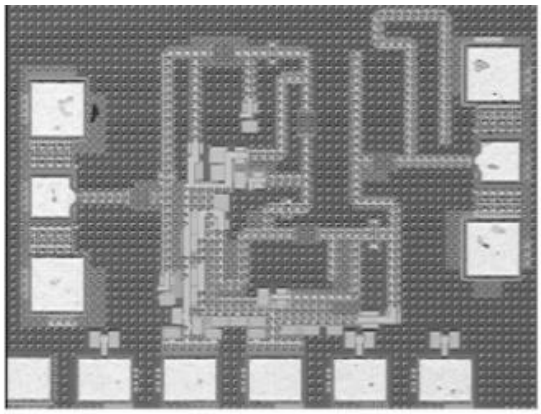


Figure 4.12. Die photo of four-stage power amplifier.

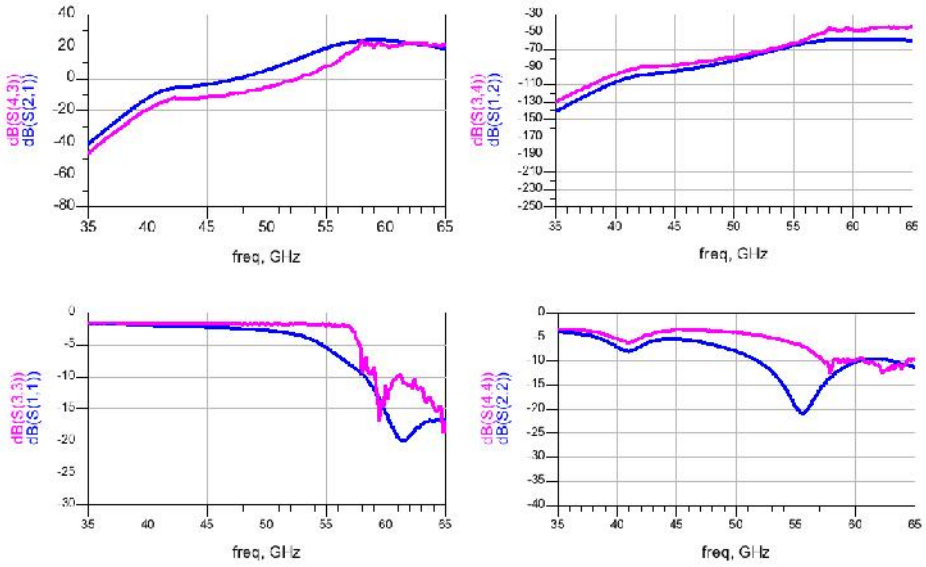


Figure 4.13. Measured and simulated S-Parameters of four-stage power amplifier.

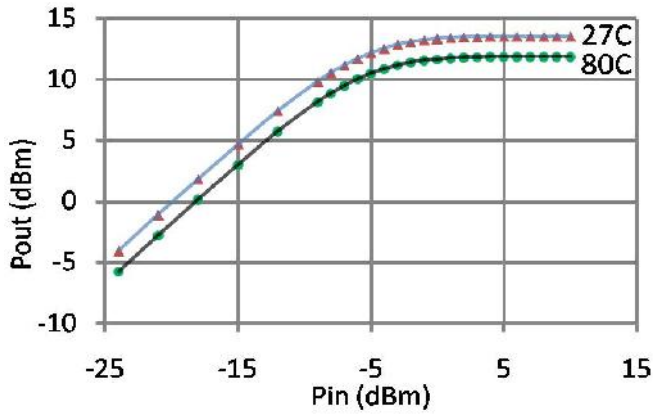


Figure 4.14. Large signal power performance of four-stage power amplifier at 60GHz.

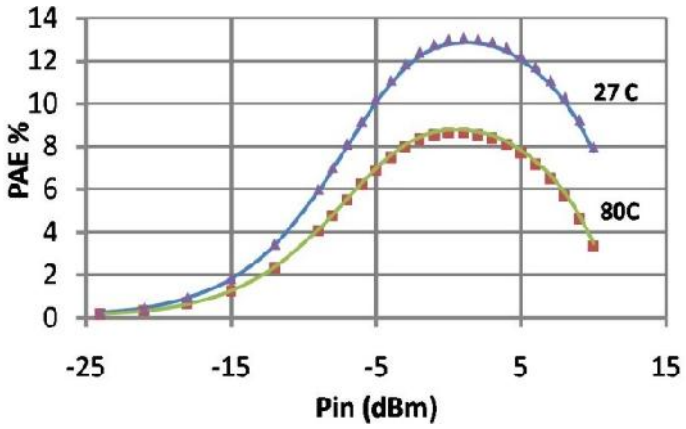


Figure 4.15. Power added efficiency of four-stage power amplifier at 60GHz.

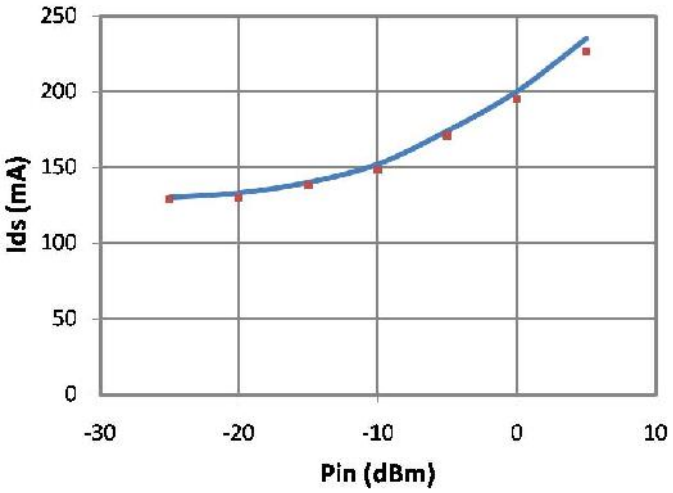


Figure 4.16. Measured and simulated drain current versus input power of four-stage power amplifier.

Table 4.2. Performance summary of four-stage 60GHz CMOS power amplifier

Gain	20dB
3dB Bandwidth	57-65GHz
Psat	13.5 dBm
P1dB	10.3 dBm
PAE	~ 13%
Chip area	0.60 mm2
Vd	1.2 V
DC Power Consumption	160 mW

As presented, the design and fabrication of 60GHz CMOS power amplifiers has provided a high degree of confidence in the developed large signal model and its practical use in millimeter-wave CMOS power amplifier design. Validation over a range of operating temperatures further validates the approach. In the next chapter, other potentially higher efficiency classes of operation will be investigated as a means of extending the performance of standard Class AB designs.

CHAPTER 5

24 GHz INVERSE CLASS F POWER AMPLIFIER DESIGN IN 130nm CMOS PROCESS

Following the use of the developed large signal model in the design of multi-stage 60GHz CMOS power amplifiers using conventional Class AB mode of operation, the next step involved experimenting with potentially higher efficiency, switch-mode type CMOS power amplifier design. Due to process limitations, as will be discussed later, an Inverse Class F single-stage power amplifier was designed using the developed model and fabricated and tested at a fundamental frequency of 24GHz. While this falls below the millimeter-wave region, it nevertheless utilizes harmonic tuning at frequencies that are well within the millimeter-wave region. Many published works such as [36]-[42] using CMOS and SiGe utilized the Class AB approach in the 18 to 24GHz range. Prior to discussion of the amplifier development, background will be given regarding the advantages, disadvantages and particular characteristics of some of the various classes of power amplifier operation.

5.1 Preliminary Considerations - Classes of Operation

The power amplifier modes of operation can be divided into switching and non-switching (or current source) modes. The latter is comprised of Class A, Class AB (employed for the 60 GHz design), Class B and Class C. The former is comprised of Class D, E, F, Inverse F as well as saturated Class A and C. Classes A, AB, B and C are best explained through the load-line and transistor I-V curves as shown in Figure 5.1. The quiescent bias point is lowered from the Vgs value at which Ids is one half of Imax in the case of Class A all the way to below transistor cutoff in the case of Class C. The corresponding drain

current waveforms for Classes A, B and C are shown in Figure 5.2. It is important to note that these waveforms are shown for the same sinusoidal input drive level applied to the gate of the FET. For Class B and Class C, as compared to Class A, it will therefore require greater input power drive to achieve the same current swing. In the case of Class B, this corresponds to 6dB greater input drive level, reducing the overall gain. However an overall advantage can be observed in the drain efficiency, given by the ratio of the output power to the DC supply power. For most power amplifier applications, however, power-added efficiency, or PAE, gives a better picture of the overall efficiency of the power amplifier by taking into account the input RF power. For lower to moderate gain amplifiers and devices, the PAE will be lower than drain efficiency, however the difference between the two efficiency metrics will decrease significantly with high amplifier gain (generally for gain values approaching 30dB or higher).

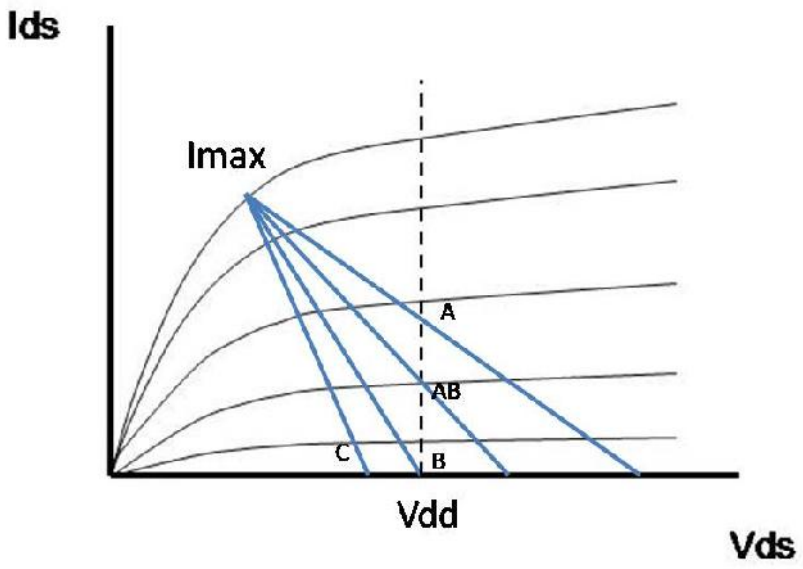
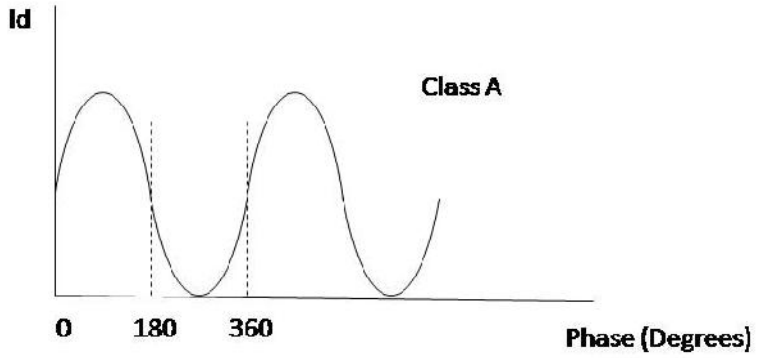
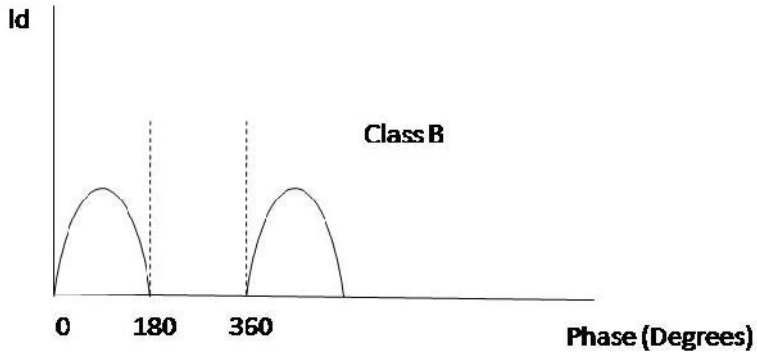


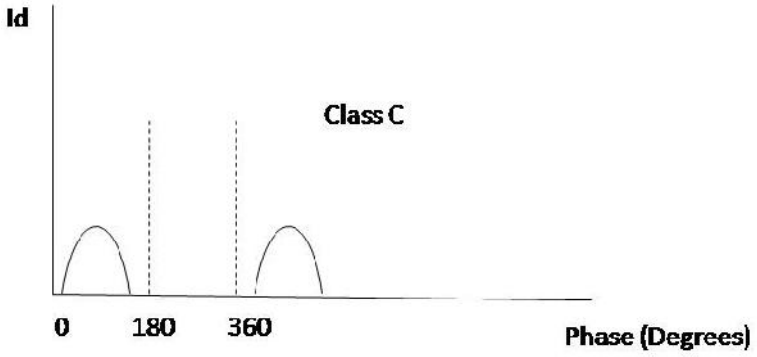
Figure 5.1. Load-lines for non switching mode classes of power amplifier operation.



(A)



(B)



(C)

Figure 5.2. Drain current waveforms for Class A, Class B and Class C modes for the same input drive.

Due to a reduction in the conduction angle as observed in Figure 5.2, reductions in the average drain current lead to reductions in the supplied DC power, which boosts the drain efficiency. For example, in the case of Class B, a maximum value of approximately 78.5 % can be achieved versus a maximum value of 50% for a Class A power amplifier. The tradeoffs of course include reduced linearity. Rather than resorting only to conduction angle reduction as in the case of Class B and C, the focus in this work is to utilize the previously developed Class AB stage and incorporate harmonic tuning to increase the overall efficiency. This leads to examining Class D, E, F and its dual, Inverse F.

Class D amplifiers utilize a series resonant circuit across a two-way switch to achieve, in the ideal case, a square voltage waveform and half sinusoid current waveform at the drain node of the FET, resulting in no overlap and theoretically, 100% drain efficiency. However, when trying to realize such an amplifier at high RF, microwave and millimeter-wave frequencies, device parasitics will start to degrade an attempt at such a realization.

Class E power amplifiers operate as a FET which behaves like a switch and is shunted by a capacitor. The current flow alternatively flows through the switch and through the parallel capacitor resulting in the characteristic Class E current and voltage waveforms.

Class F and Inverse Class F power amplifiers both require a fundamental load impedance to be presented similar to that provided to a Class AB or Class B amplifier, but with the addition of harmonic tuning to shape the drain current and voltage waveforms. The primary objective of this waveform shaping is to (as in any switch mode power amplifier) reduce the overlap between the drain current and voltage waveforms, thereby reducing the power dissipation through the transistor, thereby maximizing the

conversion of DC supplied power to RF power at the amplifier output. This is the consequence of power conservation:

$$P_{diss} = P_{DC} - \left(P_{RF,out} - P_{RF,in}\right) \qquad (5.1)$$

In the case of a Class F PA, the drain voltage waveform is squared by introducing close to an open circuit at the odd harmonic frequencies to the output of the transistor, and the drain current waveform will be a half sinusoid by introducing close to a short circuit at the even harmonics. The opposite is done in an Inverse Class F design – close to an open circuit at the even harmonic frequencies is presented to the output of the transistor, resulting in a squaring of the drain current waveform. The drain voltage waveform is now a half sinusoid. In reality, and especially at higher microwave and millimeter-wave frequencies, waveform shaping can be obtained by utilizing harmonic tuning up to the third harmonic at most due to substantial reduction of the device gain at these harmonic frequencies. In the case of a 60GHz power amplifier utilizing 90nm NMOS transistors for instance, the Ft limitations render a pure Class F or Inverse F design not even possible. The actual squaring of the voltage or current waveforms is never actually achieved in practice as there is the transistor knee voltage characteristic to take into account as well as the fact that it would require high impedance to be presented to the transistor output at an infinite number of harmonic frequencies. Nevertheless, the ideal drain current and voltage waveforms for Class F and Inverse F are shown in Figure 5.3.

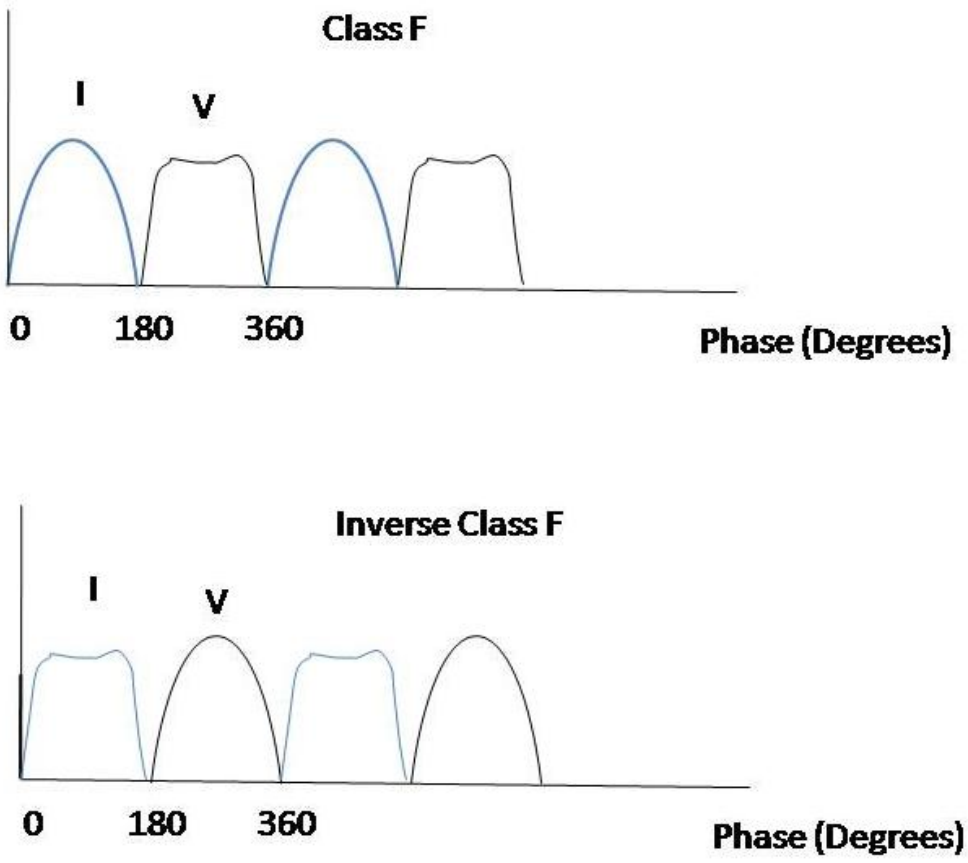


Figure 5.3. Ideal drain current and voltage waveforms for Class F and F inverse power amplifiers.

A schematic of a basic, lumped element-based Class F PA with L-C resonators is shown in Figure 5.4. There are variations of the Class F (and inverse F) PAs with tuning at one or more of the odd or even harmonics. In this case, there is tuning by means of the parallel L-C circuit at the third harmonic frequency (three times the fundamental frequency). This will serve to square the voltage waveform at the drain node of the FET.

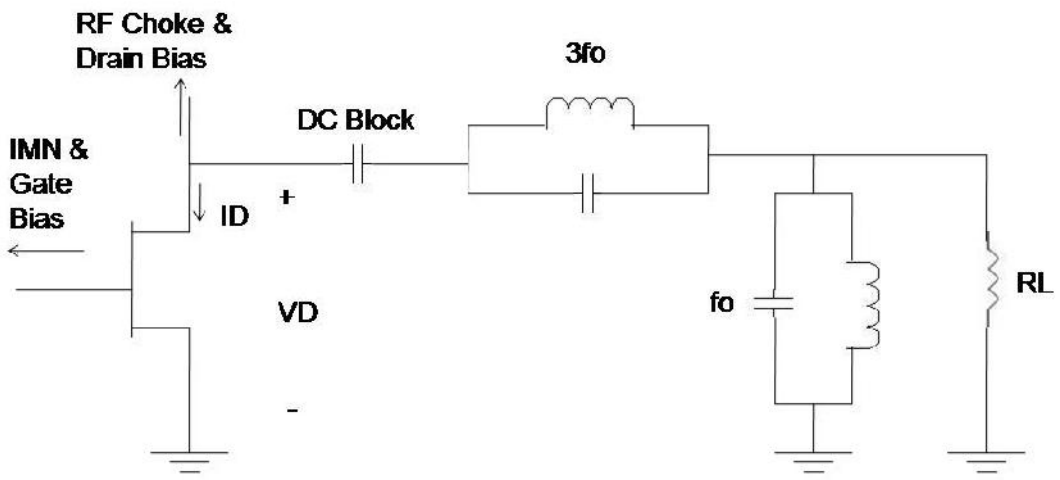


Figure 5.4. Ideal lumped element-based tuned Class F PA output matching network.

At microwave and millimeter-wave frequencies however, such harmonically-tuned matching network are typically implemented using transmission lines. A summary of the performance metrics for the classes of power amplifiers is given in Table 5.1.

Table 5.1. Summary of Tradeoffs between Classes of Operation

Class	Maximum Possible Drain Efficiency	Peak Output Power	Gain	Linearity
A	50%	moderate	large	good
B	78.5%	moderate	moderate	moderate
C	100%	small	small	poor
D	100%	large	small	poor
E	100%	large	small	poor
F/Inv. F	100%	large	small	poor

Given the various tradeoffs between the classes and the desire for higher PA efficiency, focus was placed on either designing a Class F or Inverse Class F PA. Ultimately the decision was made to design an Inverse Class F PA for a couple of reasons. While the simulated PAE and output power were comparable for the two classes, when doing harmonic load-pull simulations using the device model, in the actual output matching network design, it was easier to simultaneously provide high, low and maximum power load impedances at the second, third and fundamental frequencies respectively. Secondly, given the already poor expected linearity of the amplifier in either class, the effect of the voltage waveform on the linearity was taken into consideration. In the case of an Inverse Class F PA, the peaking of the voltage waveform at the drain of the FET, as opposed to the current waveform, will cause the instantaneous Vds to increase for a longer period of the cycle, thereby decreasing the value of Vgd and allowing for less variation of the gate-to-drain capacitance, Cgd, allowing for better linearity. The Cgd vs. Vds curves are shown in Chapter 3. These two factors formed the basis for choosing to design an inverse Class F PA in a CMOS process.

5.2 Use of 130nm BiCMOS 8HP Process

The process available for design and fabrication of the Inverse Class F PA was a 130nm (gate length) BiCMOS 8HP process from IBM. This longer-gate length process actually has Ft/Fmax comparable to the 90nm CMOS process used in the 60GHz PA designs. Although the ideal next step would be to implement a harmonically-tuned PA in a shorter gate-length process such as 45nm CMOS, this requires funding resources which were not available. Using the available process, with an Ft comparable to the ST Microelectronics process, a constraint is placed on the maximum third harmonic frequency that can be used for tuning purposes. Using this process, for sufficient third harmonic gain, the third

harmonic frequency should be less than 80GHz. So, a fundamental frequency of 24 GHz was selected for this PA, with second and third harmonics at 48 and 72 GHz.

After the frequency selection, the next step is to determine the suitability of the process for this application and the adaptability of the previously-developed device model to this process. The IBM 8HP process has, as a default, seven metal layers. In order from bottom to top-most metal layer, they are M1, M2, M3, M4, MQ, LY and AM. A cross-sectional view of the metal stack is shown in Figure 5.5. This schematic excludes the polymide and oxide layers and the polysilicon layer and substrate. It is simply used to illustrate the metal stack and contact connections between layers used in the layout of the chip.

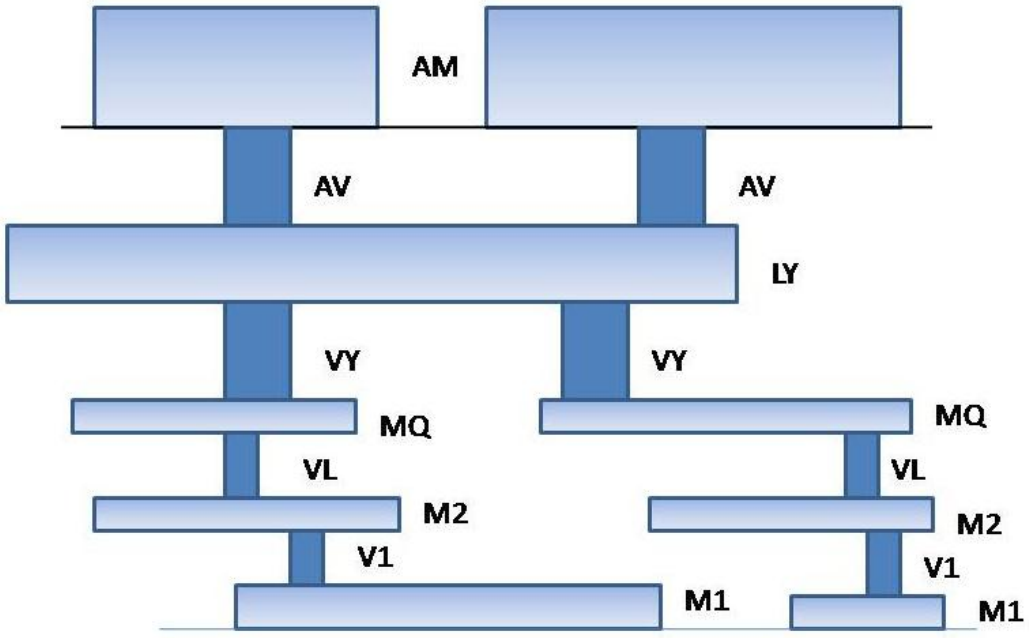
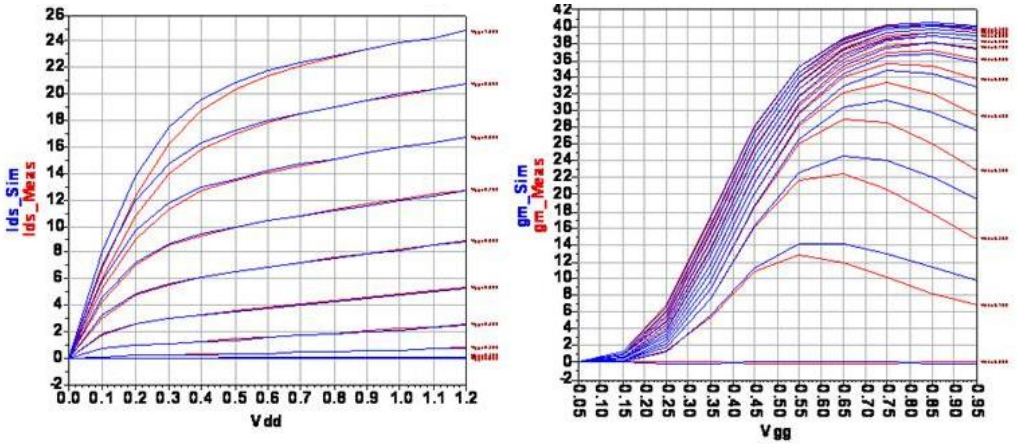


Figure 5.5. Metal stack for IBM 8HP process with only metal layers M1 through AM and contact vias shown.

Prior to starting the design process, the previously-developed large-signal model was modified and optimized to fit the DC and small signal characteristics obtained from the IBM 8HP process manual. Following this model optimization, large-signal simulations were conducted. Figures 5.6 and 5.7 show the data sheet and simulated values for Ids, transconductance and S-parameters for a 40um FET in the 8HP process.



(a) (b)

Figure 5.6. (a) DC I-V curves and (b) transconductance curves. Data Sheet (red) and simulated (blue) curves are both plotted for a 40µm device.

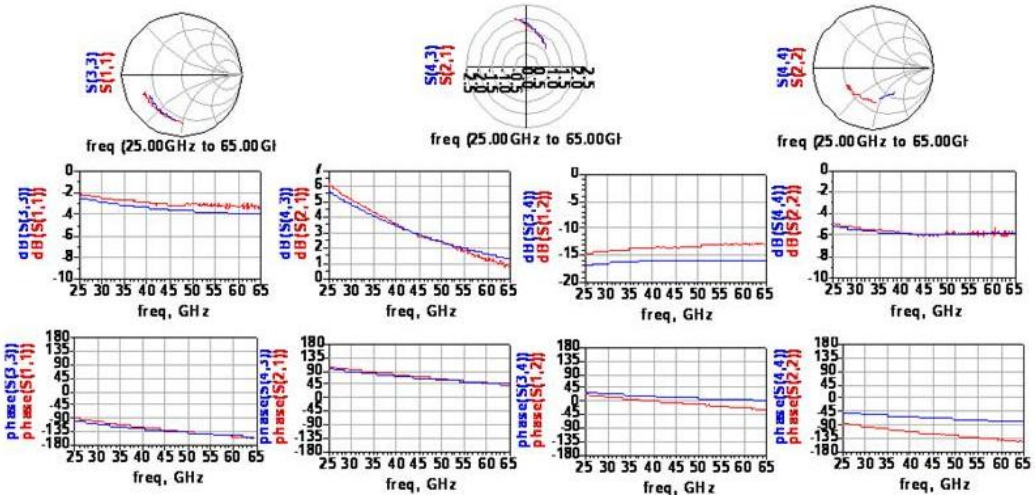


Figure 5.7. S-Parameters for a 40µm device. Data Sheet (red) and simulated (blue) curves are both plotted.

5.3 Design Approach

The first step in the amplifier design involved performing load and source pull simulations on the modified large-signal model. The device characteristics shown in Figures 5.6 and 5.7 are for a 40μm FET, however the periphery scalable model was used to simulate the desired 80μm FET, by adjusting the width parameter in the model. An 80 μm device was chosen for a saturated output power of approximately 9dBm. In Agilent ADS, a Harmonic Load and Source Pull bench was set up for model simulations to determine the load targets at the fundamental and harmonic frequencies. In Chapter 4, Load Pull simulation was performed only at the fundamental frequency. In this case the only variations are to set the second harmonic load impedance (at 48GHz) to a short (real part of the impedance close to zero as possible) and set the third harmonic load impedance (at 72 GHz) to at least three times the fundamental load impedance. This was found to be sufficient for simulating Inverse Class F operation. The fundamental load impedance was set to the optimum load impedance for maximum output power , or Zopt, for this 80 μm device. This impedance was not very different from the Zopt for the 80 μm FET from the ST Microelectronics process used in the 60GHz designs. The source impedance was also set in the load pull simulation. As before, this was determined from the input impedance looking into the FET with Zopt presented to the drain node of the device. The Smith Chart in Figure 5.8 comes from the ADS harmonic load pull simulation. Note that while the optimum load impedances for output power and PAE are shown only at the fundamental frequency, in the simulation bench setup, the second and third harmonic impedances are set to their Inverse Class F values. All of the impedances at the other harmonics are set to the default, 50Ω. The improvement in PAE as well as maximum output power are evident when compared to the maximum PAE of around 40% obtained for the same periphery device at 24GHz.

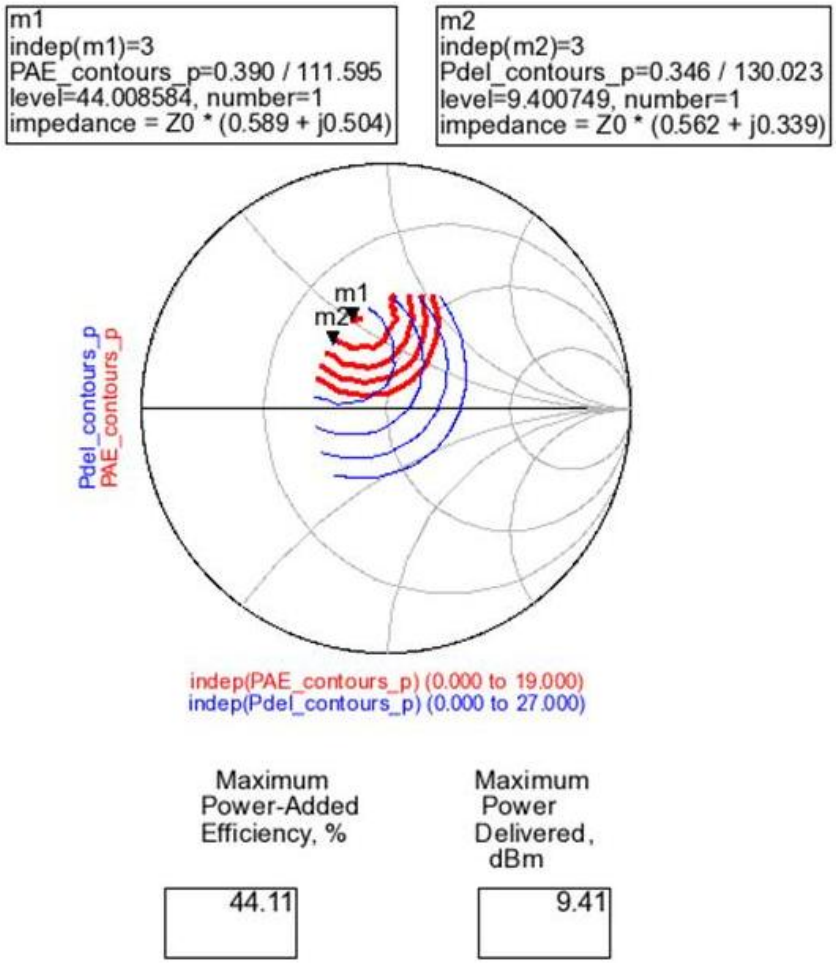


Figure 5.8. Load Pull simulation results for 80µm model with harmonic load impedances set to Inverse Class F values.

This simulation is useful as a starting point for the design, however the output power and PAE values shown are not realistically possible in an actual design, because the matching networks will incur losses in the process of transforming the 50Ω on the input and output of the amplifier to the load and source impedance which the device needs to be presented.

As a check that the harmonic load pull simulation is actually based on the desired class of power amplifier operation, it is always necessary to simulate the drain voltage and current waveforms. If indeed the harmonic impedances are correctly set and the device large signal model is accurate, then the simulated waveforms should conform to expectations. Otherwise an iterative process will need to ensue until this is the case. Figures 5.9, 5.10 and 5.11 show these waveforms (drain current waveforms on the left and drain voltage waveforms on the right) for the 80µm device for Class AB, Inverse Class F and Class F classes respectively. The expected waveform simulations for each of these three classes, indicates large signal model robustness.

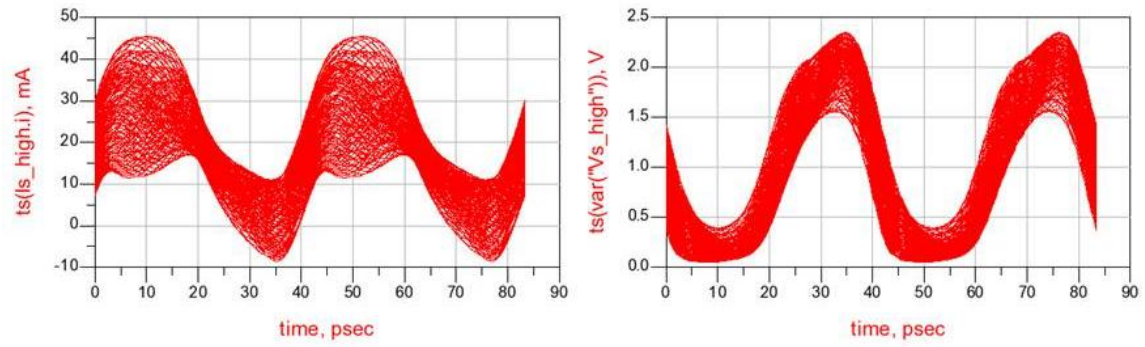


Figure 5.9. Simulated Class AB current and voltage waveforms using 80µm device model.

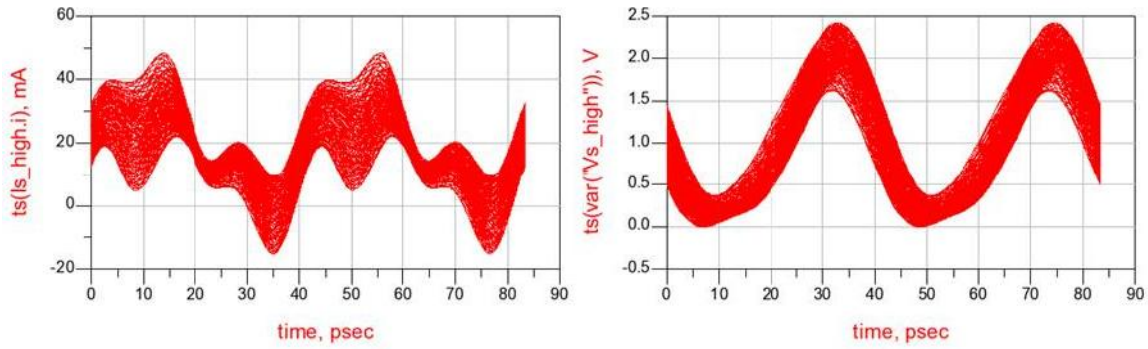


Figure 5.10. Simulated Inverse Class F current and voltage waveforms using 80µm device model.

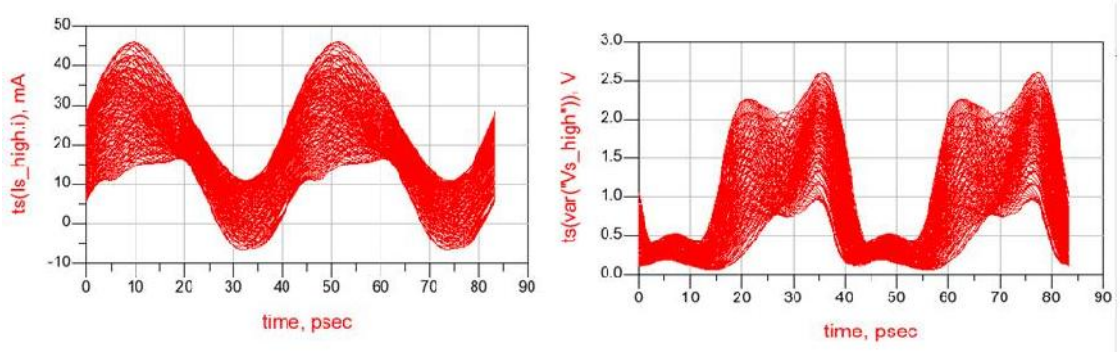


Figure 5.11. Simulated Class F current and voltage waveforms using 80µm device model.

Note that there are multiple drain current and voltage waveforms that are shown in each plot. This is due to the fact that there are multiple simulations being performed over the course of the load pull, with multiple load impedances being presented to the output of the device at the fundamental frequency over the course of the entire simulation. The desired load impedances at the fundamental frequency and second and third harmonics are shown in the Smith Chart in Figure 5.12 along with the desired source impedance for this FET.

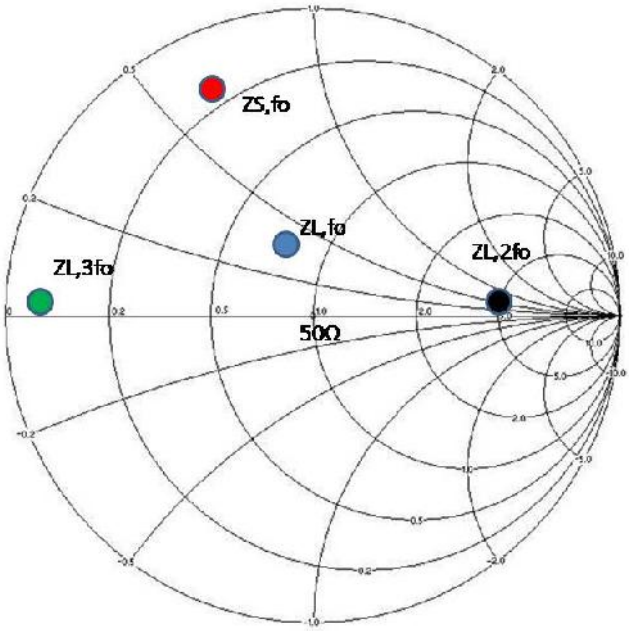


Figure 5.12. Smith Chart with load and source targets for 24GHz Inverse Class F PA.

To transform the 50Ω output to these load impedances, there are various methods that have been published. An overview of the configuration of the output matching network used in this approach is shown in Figure 5.13.

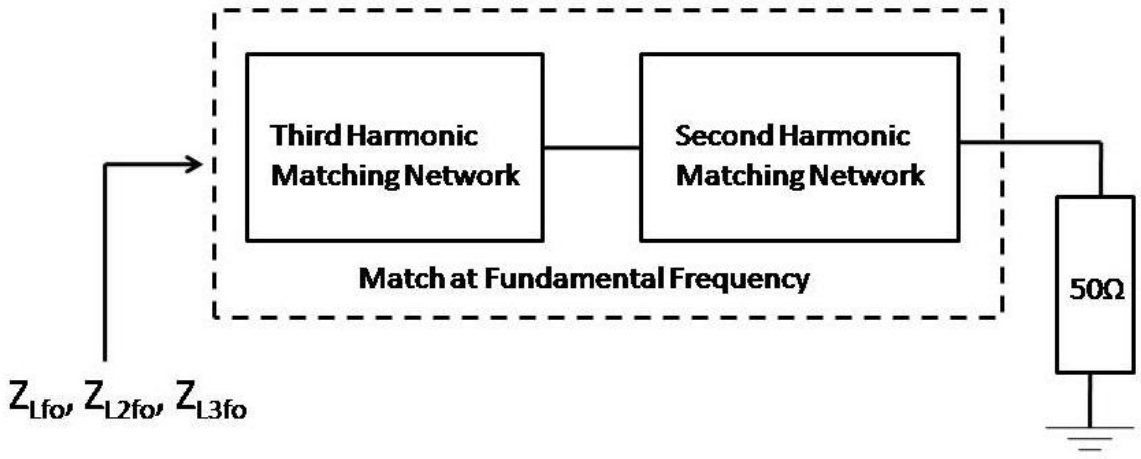


Figure 5.13. Overview of output matching network for 24GHz Inverse Class F PA.

As shown in this diagram, the design has separate matching sections for second and third harmonic matching. These sections, together, are used to simultaneously provide a match at 24GHz. The second harmonic matching network is comprised of a series microstrip line followed by a shunt microstrip line. The third harmonic network simply has a short-circuited microstrip stub. The series line length and any of the stub lengths can be adjusted to provide less of a match at either of the harmonics to provide a better match at the fundamental, if necessary. Starting with the assumption that close to an open impedance at the second harmonic can be achieved, the second harmonic network utilizes a series $\lambda/4$ line at the second harmonic which is effectively terminated by a shunt $\lambda/2$ short stub at the second harmonic. These transmission line lengths (in fractions of a

wavelength) at any given frequency are determined from the equation for lossless transmission line input impedance as shown in (5.2).

$$Z_{in}(l) = Z_0 \frac{Z_L + j \cdot Z_0 \tan(\beta l)}{Z_0 + j \cdot Z_L \tan(\beta l)} \qquad (5.2)$$

For a short-circuited stub, this reduces to the following expression:

$$Z_{in,sc}(l) = j \cdot Z_0 \tan(\beta l) \quad (5.3)$$

Given that β is equal to $2\pi/\lambda$, to achieve an open circuit, the stub length, l, must be equal to $\lambda/4$ at a given frequency. Conversely, to achieve a short circuit at a given frequency, the stub length should be a half-wavelength, $\lambda/2$.

In the second harmonic network, the shunt $\lambda/2$ short-circuit stub at the second harmonic therefore presents a short impedance to the series $\lambda/4$ line. Now that the $\lambda/4$ line is effectively terminated by a short, it is a shorted series $\lambda/4$ line. The input impedance looking into this is an open, ideally. This is how an ideally open impedance can be presented at the second harmonic to the transistor drain node. Referring everything back to the fundamental frequency, these lines become a series $\lambda/8$ line followed by a shunt $\lambda/4$ short circuit stub at the fundamental frequency.

Similarly, to provide a short impedance at the third harmonic, a shunt $\lambda/2$ short circuit stub at the third harmonic is placed before the second harmonic matching section. At the fundamental frequency, this short circuit stub has a length of $\lambda/6$. All of the transmission lines mentioned have characteristic impedance, Z_0, of 50Ω. The output matching network comprised of these ideal, lossless lines is shown in Figure 5.14. This represents a starting point for the Inverse Class F design.

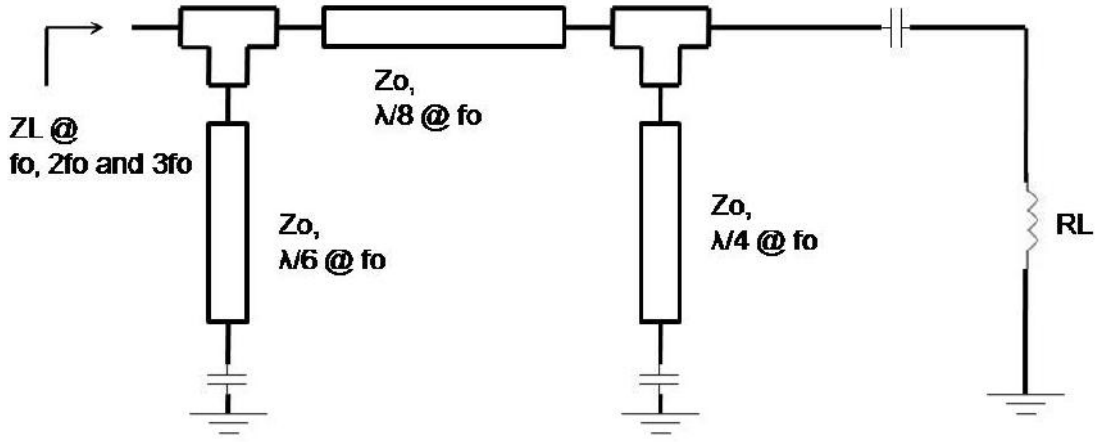


Figure 5.14. Output matching network with ideal transmission lines for 24GHz Inverse Class F PA.

The next step following the selection of matching network topology and transmission line lengths was to use LineCalc, a transmission line solver built into the Agilent ADS software, for the purpose of converting wavelengths at 24GHz to actual dimensions of microstrip lines in microns. The microstrip line width was chosen to be 8μm, determined from LineCalc, corresponding to a characteristic impedance of 50Ω. Following the initial calculation of these values, they were optimized manually to improve the power match at 24 GHz, while maintaining close to an open and short at 48 and 72 GHz respectively. When realizing the design with actual microstrip lines, it becomes apparent that when accounting for line losses in a real design, an impedance of around 5Ω can be achieved at 72 GHz and an impedance of approximately 150Ω (roughly three times Zo) at 48 GHz. However, drain current and voltage waveforms with these impedances at the second and third harmonics still provide satisfactory Inverse Class F waveforms. The input matching network is much simpler and provides impedance transformation from 50Ω to the

optimum source impedance for gain and input match at 24GHz. The approximate source impedance is shown in Figure 5.12.

The final circuit including the input matching network and the output matching network with optimization for load impedances at 24GHz as well as at 48 and 72GHz is shown in Figure 5.15. The drain bias is fed through the shunt microstrip line closest to the drain node of the FET. As is symbolically shown in the figure, an RF ground is provided for the RF signal and then the connection to the Vdd supply is made. The bias feeds for both gate and drain will be shown in greater detail following further discussion of the simulations.

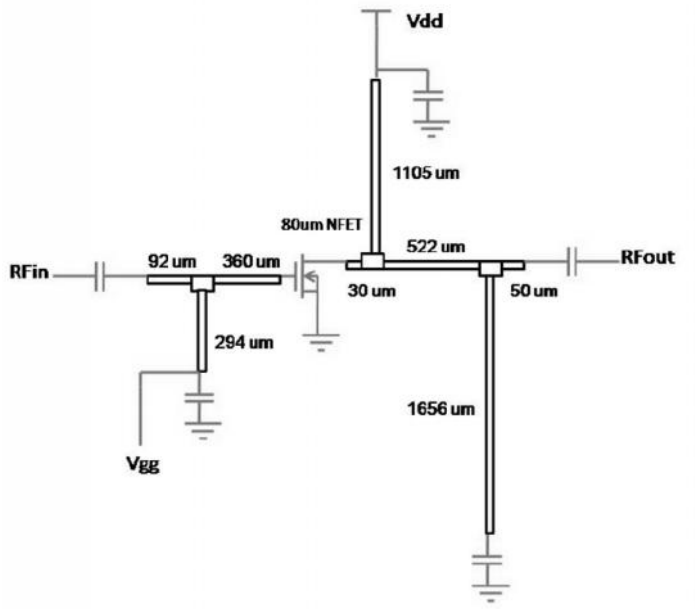


Figure 5.15. 24GHz Inverse Class F schematic with physical lengths of the microstrip lines shown.

With this implementation, the simulated S11 looking into the output matching network for frequencies up to 100GHz is shown in Figure 5.16. The matching network input impedances at 24, 48 and 72 GHz are indicated by markers and their values are shown.

As can be seen, following optimization, the optimum load impedance at 24GHz is achieved along with harmonic impedances close to their Inverse Class F targets.

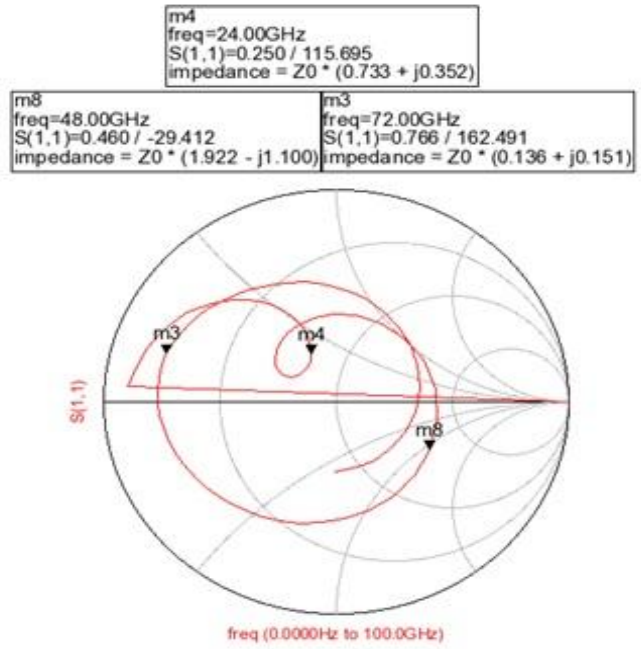


Figure 5.16. Smith Chart with realized output matching network impedances – ADS simulation.

The corresponding drain current and voltage waveforms are shown in Figure 5.17. It is important to note that there are multiple waveforms shown because the ADS large signal simulation is swept over a range of power levels.

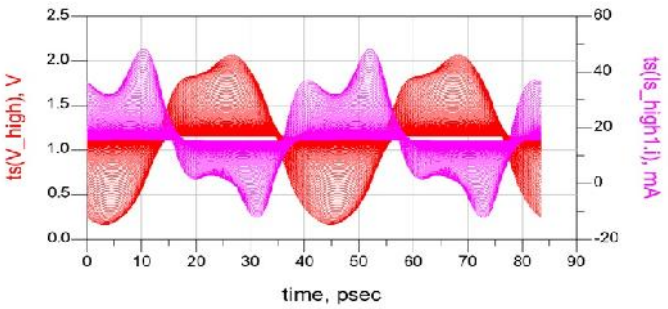


Figure 5.17. Drain current and voltage waveforms for simulated Inverse Class F PA.

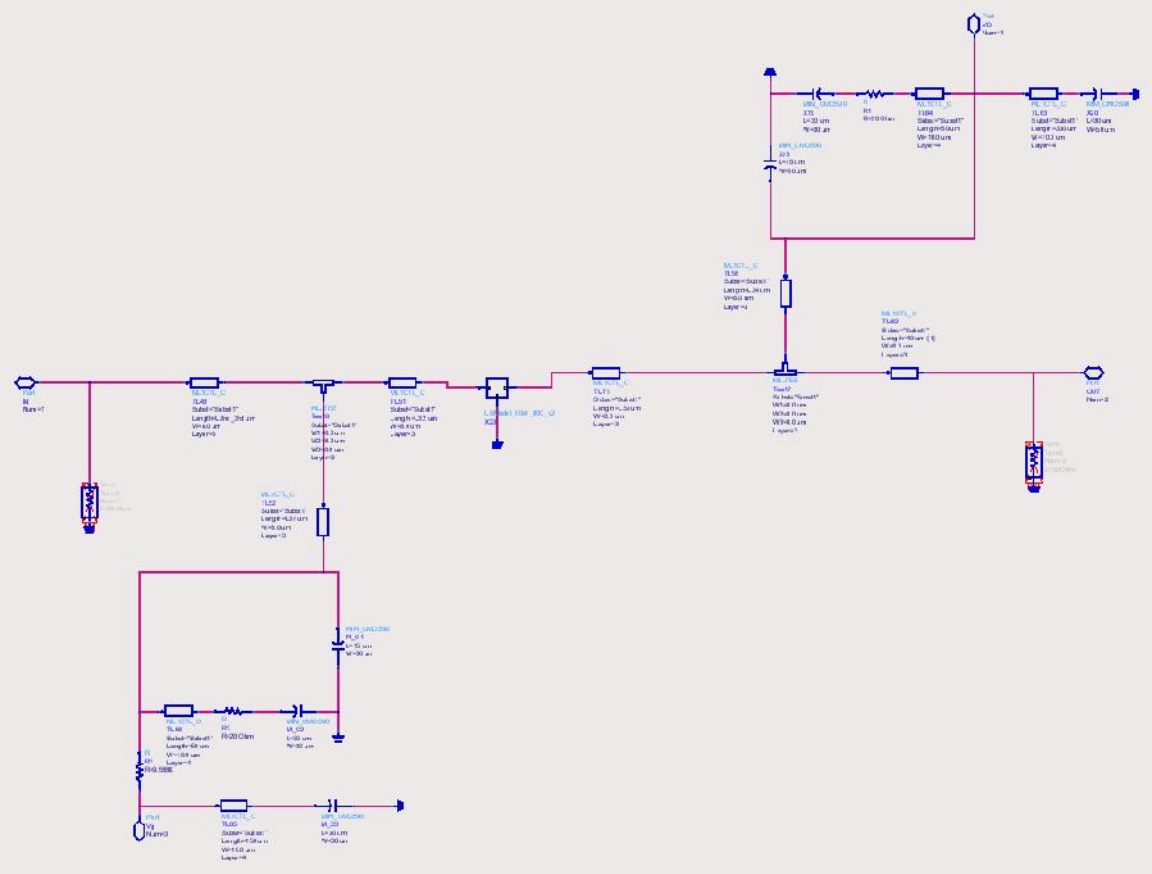


Figure 5.18. ADS Schematic of PA with gate and drain bias feed networks shown.

The gate and drain bias voltages are applied through DC bias pads to the FET. The gate current is negligible for these devices, however the drain current flows through the shunt transmission line and then from drain to source in the FET. The DC blocks are actually in a higher level of the schematic hierarchy and that is why they are not visible in the schematic in Figure 5.18. The shunt microstrip lines used for bias feed have a MIM capacitor to ground, forming an RF ground. Also seen in this schematic are shunt bypass capacitors, which are used to eliminate some of the unwanted low frequency gain spikes that can cause instability. For very low frequencies, such as 100MHz, it is necessary to use very large capacitor values on the order of 100pF or higher, which are typically used

off-chip. However on-chip it is possible to use smaller valued MIM capacitors and eliminate spurious gain at frequencies that are higher than this but are still well below the frequency band of interest. Lastly, there is a large shunt resistor, on the order of 1 to 2 KΩ placed in the gate bias feed. One of the reasons for including this resistor is to increase the isolation between the gate bias feed and the RF signal path. It is also useful in the event that if the FET ever begins oscillating and hence has a negative input resistance, the total input resistance will remain positive and prevent bias circuit oscillation [43].

Although the small and large signal power amplifier performance will be shown in section 5.5 along with the measured results, a highly important simulation for unconditional stability will be shown here. In Figure 5.19, the simulated K-factor of the PA is shown over the frequency range spanning DC to 70GHz. Along with ensuring that the amplifier S11 and S22 always have magnitude less than 0dB, it is necessary to ensure that the K-factor is greater than unity over a very broad frequency range. There are of course, other additional stability criteria that should be met, however this is a necessary starting point.

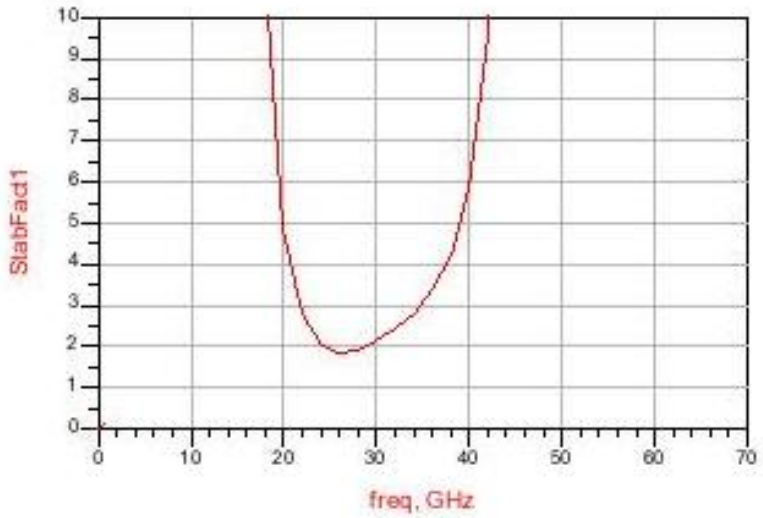


Figure 5.19. K-factor simulation of 24GHz Inverse Class F PA.

The K-factor is given by the following expression involving all four two-port S-parameters:

$$K = \frac{1 - |S11|^2 - |S22|^2 + |S11 \cdot S22 - S21 \cdot S12|^2}{2|S21 \cdot S12|} \quad (5.4)$$

5.4 Layout Approach

The inverse Class F power amplifier layout is shown in Figure 5.20. The layout utilized the metal layers, capacitors, pads and active devices available in the IBM 8HP design kit.

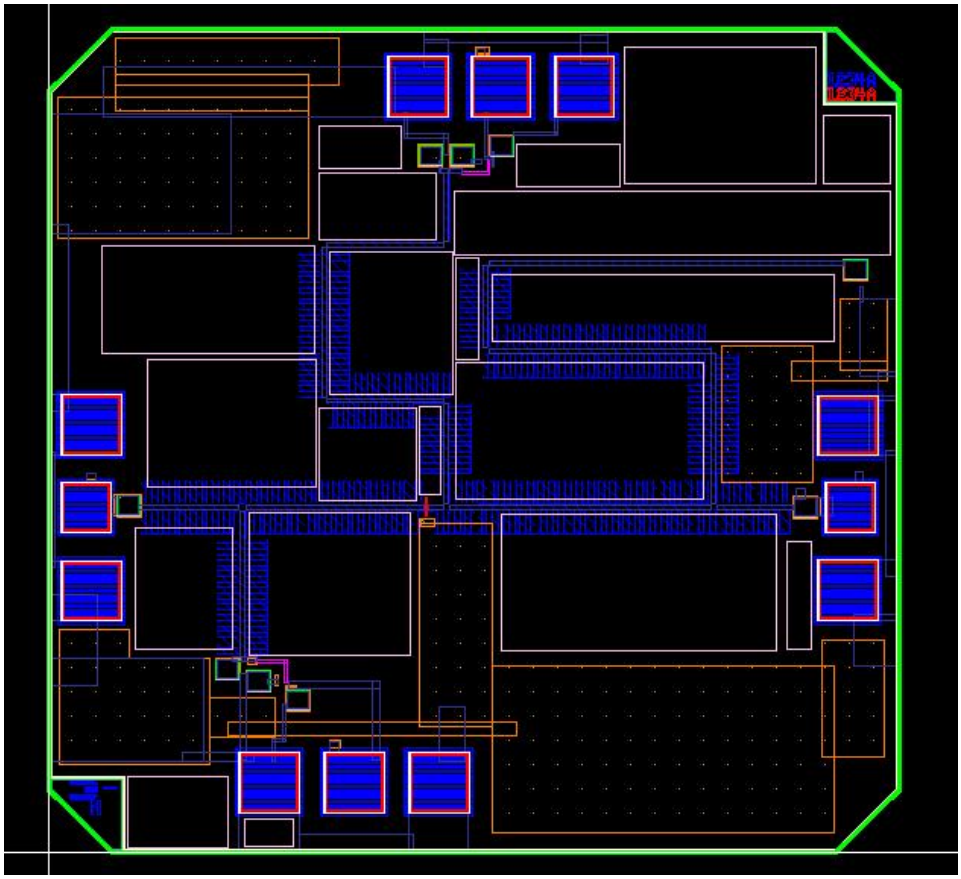


Figure 5.20. Cadence layout of 24GHz Inverse Class F PA.

An 80µm NMOS transistor, designated as an NFET in this particular design kit, was chosen to achieve a saturated output power of 9dBm in simulation. The microstrip lines in the matching network sections were implemented using the AM metal layer for the RF signal layer and M1 for ground. Ground-signal-ground pads were taken directly from the 8HP library with connections from the RF signal pads made directly to DC-blocking MIM capacitors with values of 1.8 pF. On the gate and drain bias side, bypass capacitors were also implemented with MIM capacitors and additional resistors were chosen to be oppcres-type polysilicon resistors. These resistors were primarily used on the gate side of the device as discussed in the previous sub-section.

The NFET used in this process had 80 gate fingers, each of which was 1µm wide. The gate length, L, for all gate fingers was the process default, 0.13µm. A gate feed was added to the gate side of the NFET in the polysilicon layer, connecting all of the gate fingers together and then by using vias, was connected to the AM microstrip line in the input matching network. Similarly, all of the drain fingers were connected in the M2 layer and then using vias, was connected to the AM microstrip line in the output matching network. Furthermore, a source ring in the M1 layer was inserted around the periphery of the NFET, connecting to the source contacts on both side of the transistor.

Also visible in the Cadence layout are large rectangles of the thick top metal layers that fill the areas in between the transmission lines, passive devices and pads. These are the AM, LY and MQ metal layers that exist within the actual circuit, but must be added throughout the remaining chip area as well to meet the metal-fill density requirements of the foundry. Following completion of the layout, design rule checks (DRC) and layout versus schematic (LVS) were performed and when finally clean, the layout file was submitted to IBM for fabrication through the MOSIS service.

5.5 Measured and Simulated PA Characteristics

A die photo of the fabricated chip is shown in Figure 5.21. The total chip dimensions are 1.6x1.5mm. Shown in Figures 5.22, 5.23, 5.24 and 5.25 are the measured and simulated S-Parameter magnitudes for the 24GHz Inverse Class F power amplifier. As observed, there is reasonable agreement between measurements and the performance which the modified model predicts. In addition, in Figures 5.26 and 5.27, the simulated and measured large signal power performance and power added efficiency are plotted, again showing reasonable agreement. It should be observed however, that the efficiency values obtained in both simulation and measurement are lower than what one would expect from a power amplifier designed using a switch-mode topology. This can be primarily attributed to the relatively low gain available in the CMOS process used, especially at the frequency of interest.

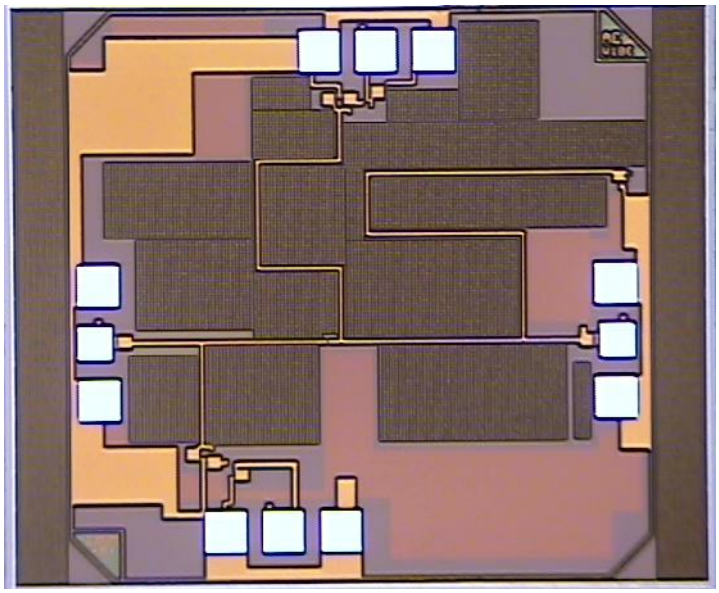


Figure 5.21. Die Photo of 24GHz Inverse Class F PA.

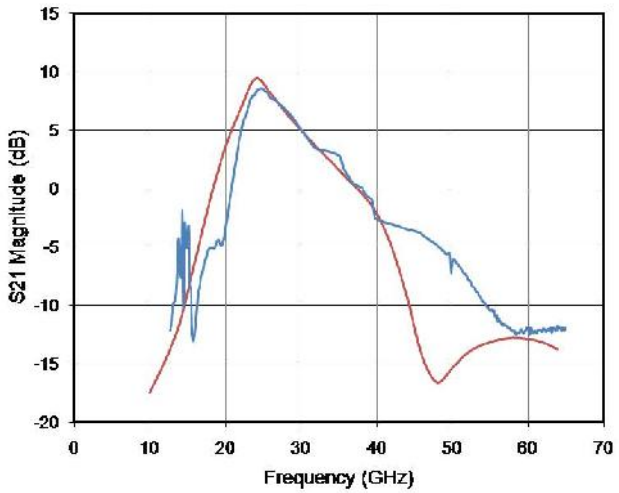


Figure 5.22. Measured (Blue) and Simulated (Red) S21 Magnitude of 24GHz Inverse Class F PA. Vdd=1.2V and Vgs=0.5V.

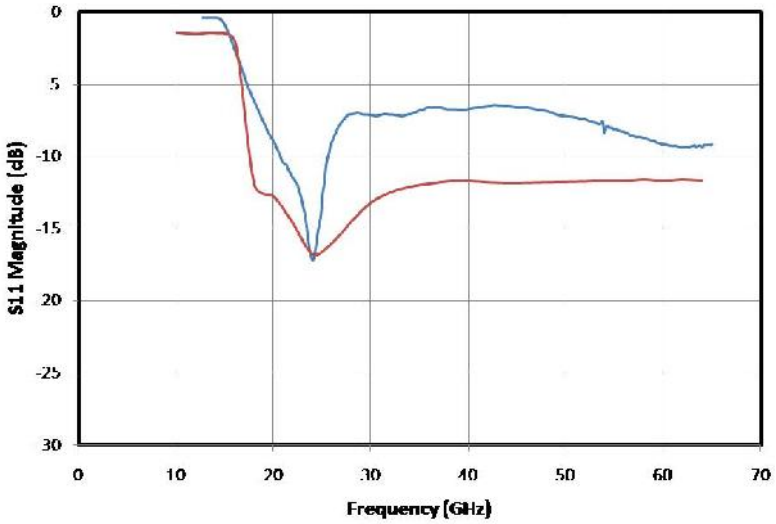


Figure 5.23. Measured (Blue) and Simulated (Red) S11 Magnitude of 24GHz Inverse Class F PA. Vdd=1.2V and Vgs=0.5V.

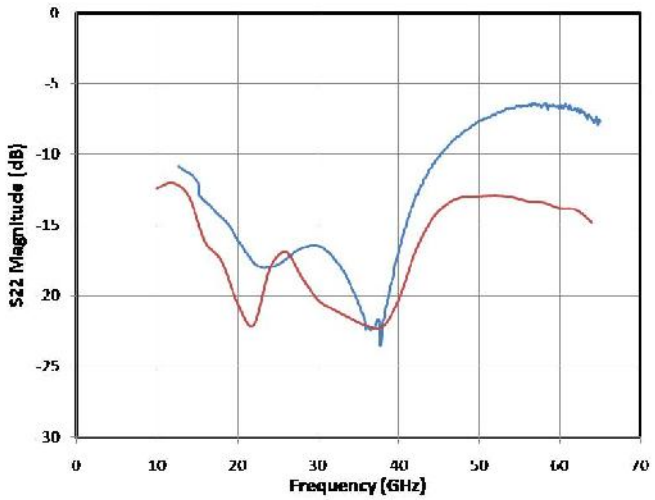


Figure 5.24. Measured (Blue) and Simulated (Red) S22 Magnitude of 24GHz Inverse Class F PA. Vdd=1.2V and Vgs=0.5V.

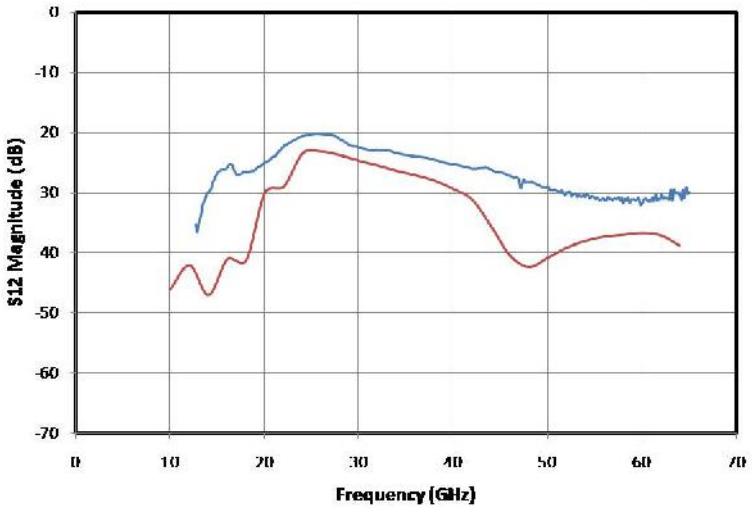


Figure 5.25. Measured (Blue) and Simulated (Red) S12 Magnitude of 24GHz Inverse Class F PA. Vdd=1.2V and Vgs=0.5V.

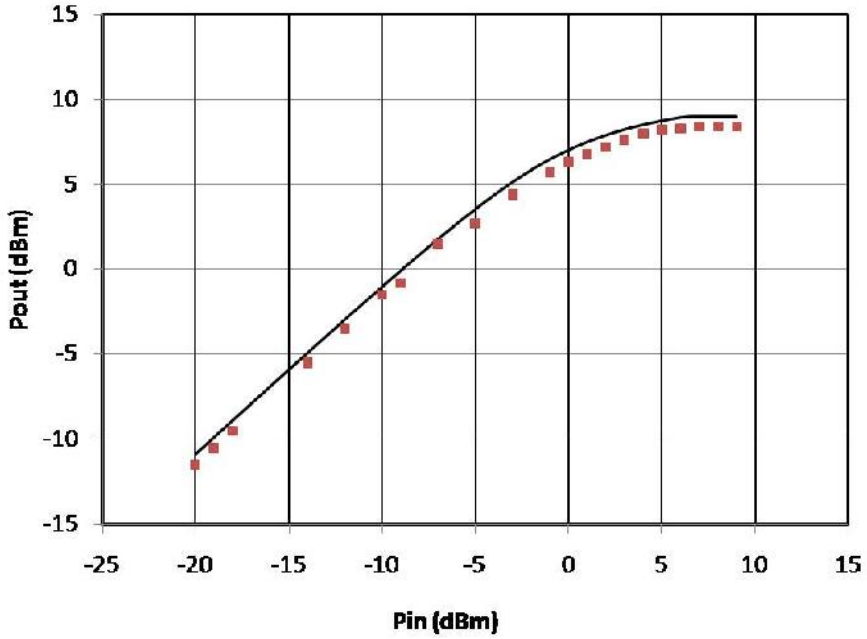


Figure 5.26. Measured (Dots) and Simulated Large Signal Power sweeps of 24GHz Inverse Class F PA. Vdd=1.2V, Vgs=0.5V and quiescent Ids=14mA.

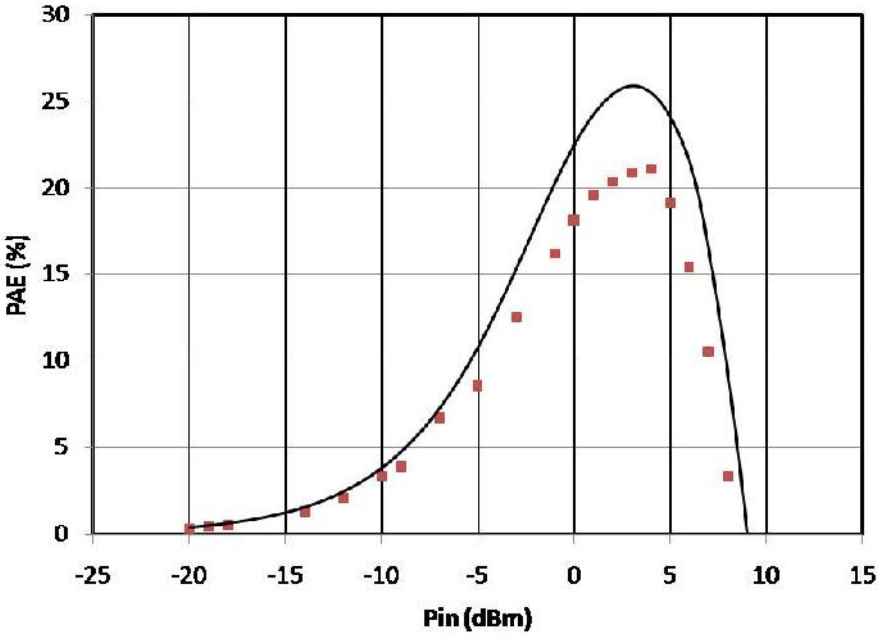


Figure 5.27. Measured (Dots) and Simulated (Solid Curve) Power-added Efficiency of 24GHz Inverse Class F PA. Vdd=1.2V, Vgs=0.5V and quiescent Ids=14mA.

The measured and simulated performance validate this method of extending Class F or Inverse Class F power amplifier design techniques using CMOS technology to 24 GHz, and potentially into the millimeter-wave frequency range if using a 65nm, 45nm or shorter gate length CMOS process.

CHAPTER 6

SUMMARY & CONCLUSIONS

6.1 Summary

A device modeling method optimized for millimeter-wave CMOS power amplifier design has been demonstrated both through device level verification and in power amplifier design at both 60 and 24 GHz. The latter frequency, while not in the millimeter-wave region, was chosen for the design of an Inverse Class F CMOS power amplifier due to process limitations on the fundamental frequency of a harmonically tuned power amplifier. Nevertheless, this investigation at 24 GHz demonstrates the feasibility of using a transmission line-based harmonically tuned class of power amplifier which can be extended into the millimeter-wave frequency range given a CMOS process with a shorter gate length. The use of the device model in the design of two separate 60GHz CMOS power amplifiers was also demonstrated, with the features of device periphery scaling and temperature dependency enhancing the model's predictive capabilities.

6.2 Conclusions

There are a number of conclusions learned from this research. First, the developed empirical model is more robust in large signal harmonic balance simulators such as ADS when simulating multistage CMOS power amplifiers, as opposed to models developed using macro-models such as the BSIM3. Furthermore, as is necessary in any large signal device modeling approach, the large signal model accurately narrows down to the small

signal case. This is reflected by close S-Parameter agreement over a broad frequency range (DC through 65 GHz) between model simulations and measurements of both discrete transistors and multi-stage power amplifiers. Another conclusion is that the method of incorporating temperature scalability into the drain current generator and parasitic elements proves to be robust when over-temperature measurements of multi-stage CMOS power amplifiers are compared with simulations. Furthermore, device periphery scaling is incorporated effectively, as demonstrated by the measurement and simulation agreement of 60GHz power amplifiers utilizing devices of increasing periphery in each successive amplifier stage.

The model was developed and works well for predicting the large signal fundamental frequency power characteristic and power added efficiency for power amplifiers at both 60 and 24 GHz. Work has not focused on assessing linearity, as simulated and measured by metrics such as intermodulation distortion, two-tone tests, second, third and higher order harmonic output power sweeps. Reasons for not accounting for these effects are primarily due to linearity standards not being clearly defined at the current time for the millimeter-wave wireless standards (such as W-HDMI or other standards still being developed) for which these amplifiers were developed.

Regarding the model extraction methodology, there are a few issues that should be noted. This model is more empirically-based than BSIM3 and similar industry-standard models, but has fewer parameters and is simpler, while working well for the intended application - millimeter-wave CMOS multistage power amplifier design in variable temperature environments. Of the modeling approaches presented in this dissertation, the chosen equation-based model and BSIM3, both require extensive measurements prior to model extraction, so the target application will determine the type of model to be used, as opposed to the preliminary amount of measurement effort needed. The modeling approach presented in this dissertation requires, when implemented in a commercially-

available simulator such as ADS, a set of initial model parameter values for the nominal device size and nominal temperature (27 degrees Celsius) combination. This set of values would be nominal and can be used in conjunction with the temperature and periphery scaling parameters or stand-alone. Additionally, the modeling approach assumes an applied drain voltage in the typical range (0.9-1.2V) for standard CMOS transistors and this avoids device breakdown limitations. The model is easily adaptable to other gate length processes. This is evidenced by its use in the design of a 24 GHz Inverse Class F CMOS power amplifier. This design also demonstrates that harmonically-tuned CMOS power amplifiers can be extended to 24 GHz as well as to higher frequencies in the millimeter-wave range if a shorter gate length CMOS process is available.

6.3 Contributions

- Development of empirical, equation-based large signal device model for sub-100nm CMOS processes to be used in common harmonic balance circuit simulators for multi-stage millimeter-wave CMOS power amplifier design. This approach is not an extension of previous empirical models developed for device technologies such as Gallium Arsenide-based FETs. Rather it uses novel equations tailored for these CMOS processes.

- The basic model incorporates a temperature dependent current generator over a broad range of operating bias voltages in addition to temperature-dependent parasitics. This represents a novel and more robust temperature-dependent CMOS large signal model as compared to previous CMOS modeling efforts, such as those from the BSIM family.

- The optimization of parasitic extraction of CMOS devices through 65 GHz and demonstration of this accurate extraction in both discrete device and amplifier small and large signal simulations and measurements.

- As opposed to the vast majority of CMOS power amplifier designs, the 60GHz CMOS multistage power amplifier has been designed using source and load pull simulations of a custom, empirical large signal model tailored to millimeter-wave applications. Most CMOS power amplifier designs utilized RF-enhanced digital core models such as those from the BSIM family which can have large signal convergence issues in harmonic balance simulations.

- The development of a large signal CMOS device model that is easily modifiable to other nearby (such as 130nm) CMOS process nodes as demonstrated by its use in the design of 24 GHz Inverse Class F PA.

- The design and fabrication of a 24 GHz Inverse Class F CMOS power amplifier at 24GHz. Most harmonically-tuned CMOS PAs have been demonstrated for cellular applications (i.e. around 2 GHz) and use lumped element-based matching networks and harmonic tuning. The distributed harmonic tuning/matching network presented in this amplifier demonstrates feasibility of implementation and differs from the Class AB CMOS power amplifiers typically used at these higher frequencies. Furthermore, the design approach used is unique in that it does not explicitly implement power/efficiency matching at the fundamental frequency (24 GHz) but rather first optimizes the second and third harmonic matching and uses this network to realize the fundamental frequency match.

6.4 Suggestions for Future Research

- The current generator in the device model will be extended to model the Ids-Vds characteristics in the breakdown region. This can be done by adding empirical equations with exponential terms (and with the tail extending toward lower Vds) to the existing drain current equations. This will allow for any drain voltage to be applied in simulation while accurately accounting for the breakdown voltage of the particular device.

- The model can be modified for a shorter gate length CMOS process (such as 45nm) and used in the design of a power amplifier with harmonic tuning (Class F, Inverse F or E) centered around a millimeter-wave frequency (preferably above 40 GHz). Also, a Class AB CMOS power amplifier could be designed and fabricated at the same frequency and the efficiency performance of the two could be compared.

6.5 Publications Resulting from this Work

First Author:

"Temperature-Dependent Scalable Large Signal CMOS Device Model Developed for Millimeter-Wave Power Amplifier Design", 2011 RFIC Symposium Digest.

"Temperature-Dependent Scalable Millimeter-Wave Large Signal Model for 90nm CMOS". Asia Pacific Microwave Conference, December 2008.

Co-Author:

"60 GHz CMOS Power Amplifier with 20dB Gain and 12 dBm Psat," Microwave Symposium Digest, 2009. IEEE MTT-S International, pp. 537-540.

REFERENCES

[1] H. Zirath *et al.*, "Development of 60-GHz front-end circuits for a high-data-rate communication system," *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1640–1649, Oct. 2004.

[2] C.-H. Wang *et al.*, "A 60 GHz transmitter with integrated antenna in 0.18 μm SiGe BiCMOS technology," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2006, pp. 659–668.

[3] J. Laskar, S. Pinel, D. Dawn, S. Sarkar, B. Perumana, and P. Sen, "The next wireless wave is a millimeter wave," *Microw. J.*, pp. 22–36, Aug. 2007.

[4] S. Emami, C. H. Doan, and A. M. Niknejad, "A highly integrated 60 GHz CMOS front-end receiver," in *Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2007, pp. 190–191.

[5] T. Yao, M. Q. Gordon, K. K. W. Tang, K. H. K. Yau, M.-T. Yang, P. Schvan, and S. P. Voinigescu, "Algorithmic design of CMOS LNAs and PAs for 60-GHz radio," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1044–1057, May 2007.

[6] S. P. Voinigescu *et al.*, "CMOS SOCs at 100 GHz: System architectures, device characterization, and IC design examples," in *IEEE RFIC Symp. Dig.*, 2007, pp. 1971–1974.

[7] M. Varonen *et al.*, "Millimeter-wave integrated circuits in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 9, pp. 1991–2002, Sep. 2008.

[8] S. Pinel, S. Sarkar, P. Sen, B. Perumana, D. Yeh, D. Dawn, and J. Laskar, "A 90 nm CMOS 60 GHz radio," in *Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2008, pp. 130–131.

[9] S. Sarkar, P. Sen, B. Perumana, D. Yeh, D. Dawn, S. Pinel, and J. Laskar, "60 GHz single-chip 90 nm CMOS radio with integrated signal processor," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2008, pp. 1167–1170.

[10] B. A. Floyd, S. K. Reynolds, U. R. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, "SiGe bipolar transceiver circuits operating at 60 GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156–157, Jan. 2005.

[11] S. K. Reynolds *et al.*, "A silicon 60-GHz receiver and transmitter chipset for broadband communications," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2820–2831, Dec. 2006.

[12] S. Hakusui, http://www.rfglobalnet.com/article.mvc/Fixed-Wireless-Communications-at-60GHz-Unique-0001, (Accessed January 1, 2012).

[13] U. R. Pfeiffer and D. Goren, "A 20 dBm fully-integrated 60 GHz SiGe power amplifier with automatic level control," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1455–1463, Jul. 2007.

[14] D. Dawn, S. Sarkar, P. Sen, B. Perumana, D. Yeh, S. Pinel, and J. Laskar, "17-dB-gain CMOS power amplifier at 60 GHz," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2008, pp. 859–862.

[15] A. Valdes-Garcia *et al.*, "60 GHz transmitter circuits in 65 nm CMOS," in *IEEE RFIC Symp. Dig.*, 2008, pp. 641–644.

[16] M. Tanomura, Y. Hamada, S. Kishimoto, M. Ito, N. Orihashi, K. Maruhashi, and H. Shimawaki, "TX and RX front-ends for 60 GHz band in 90 nm standard bulk CMOS," in *Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2008, pp. 558–559.

[17] *"Agilent User's Guide,"* Agilent Technol., Santa Clara, CA, 2009. [Online]. Available: http://eesof.tm.agilent.com

[18] D. M. Pozar*, Microwave Engineering*, 2nd ed. New York: Wiley, 1998.

[19] S. Emami, C.H. Doan, A.M. Niknejad, and R.W. Brodersen, "Large- Signal Millimeter-Wave CMOS Modeling BSIM3", 2004 IEEE Radio Frequency Integrated Circuits Symposium.

[20] I. Angelov, M. Ferndahl, F. Ingvarson, H. Zirath, H.O. Vickes, "CMOS Large Signal Model for CAD," 2003 IEEE MTT-S International Microwave Symposium Digest, Vol. 2, June 2003, pp. 643-646.

[21] W.R. Curtice, "GaAs MESFET Modeling and Nonlinear CAD," IEEE Transactions on Microwave Theory and Techniques, Vol. 36, No. 2, February 1988, pp. 220-230.

[22] M.J. Deen and T.A. Fjeldly *, CMOS RF Modeling, Characterization and Applications*, Singapore: World Scientific Publishing Co., Ptc., Ltd., 2002.

[23] M.D. Godfrey, "CMOS Device Modeling for Sub-threshold Circuits," IEEE Trans. on Circuits and Systems –II: Analog and Digital Signal Processing, vol. 39, No. 8, August 1992, pp. 532-539.

[24] B.G. Perumana, R. Mukhopadhyay, S. Chakraborty, C.H. Lee, and J. Laskar, "A Low-power fully monolithic sub-threshold CMOS receiver with integrated LO Generation for 2.4 GHz Wireless PAN Applications," IEEE Journal of Solid-State Circuits, Vol. 43, No. 10, October 2008, pp. 2229–2237.

[25] A. Abuelmaatti, I. Thayne, "A new RF SiCMOS SDD Model for Quantifying Individual Contribution to Distortion from Transistor's Nonlinear Parameters," 2006 IEEE International Conference on Electronic Circuits and Systems, December 2006, pp. 5-8.

[26] A. Grebennikov, RF and Microwave Power Amplifier Design, New York, NY: McGraw Hill Professional Engineering, 2004, pp. 77.

[27] G. Dambrine, A. Cappy, F. Heliodore, E. Playez, "A new method for determining the FET small-signal equivalent circuit," IEEE Transactions Microwave Theory and Techniques, Vol. 36, pp. 1151- 1159, June 1988.

[28] I. Kwon, M. Je, K. Lee, H. Shin, "A Simple and Analytical Parameter Extraction Method of Microwave MOSFET," IEEE Transactions on Microwave Theory and Techniques, Vol. 50, Issue 6, June 2002, pp. 1503-1509.

[29] I. Angelov, N. Rorsman, J. Stenarson, M. Garcia, H. Zirath, "An Empirical Table-Based FET Model," IEEE Transactions on Microwave Theory and Techniques, Vol. 47, No. 12, December 1999, pp. 2350 – 2357.

[30] K. Kundert, www.designers-guide.org/modeling/varactors.pdf, (Accessed Dec. 6, 2011).

[31] S. Hong, S. Lee, H. Choi, H. Kang, Y. Jeong, "RF Characteristics of 80nm CMOS Transistors," The 2nd International Symposium on Nanomanufacturing," November 2004, pp. 353-356.

[32] S. Pinel, S. Sarkar, P. Sen, B. Perumana, D.Yeh, D.Dawn,, J. Laskar, "60Ghz CMOS 90 nm radio", ISSCC 2008, Feb. 2008.

[33] S. Sarkar, P. Sen, B. Perumana, D.Yeh, D.Dawn, S. Pinel, J. Laskar,, "60Ghz CMOS 90 nm radio with integrated signal processor", IMS 2008, June 2008.

[34] J.M. Golio, RF Microwave Handbook, CRC press, 2001

[35] S. Cripps, *RF Power Amplifiers for Wireless Communication*, 2nd Edition, Norwood,MA: Artech House Inc., 2006

[36] N. Kinayman, A. Jenkins, D. Helms, and I. Gresham, "Design of 24 GHz SiGe HBT balanced power amplifier for system-on-a-chip ultra-wideband applications," in Radio Frequency Integrated Circuits (RFIC) Symposium Digest, June 2005.

[37] J.P. Comeau, J.M. Andrews, and J.D. Cressler, "A monolithic 24 GHz, 20 dBm, 14% PAE SiGe HBT power amplifier," in Proceeding of the European Microwave Conference. Sept. 2006, pp. 91-93.

[38] Jing-Lin, Kuo, Zuo-Min Tsa, Huei Wang, “A 19.1-dBm fully-integrated 24 GHz power amplifier using 0.18-um CMOS technology,” in Proceeding of the European Microwave Conference. Oct. 2008. pp. 558-561.

[39] H. Portela, V. Subramaniam, G. Boeck, “Fully integrated high efficiency K-band PA in 0.18-um CMOS technology,” in Proceeding of the Microwave and Optical Conference. Nov. 2009. pp. 393-396.

[40] C. Cao, H. Xu, Y. Su, and K.K. O, “An 18 GHz, 10.9 dBm fully-integrated power amplifier with 23.5% PAE in 130-nm CMOS,” in Proceeding of ESSCIRC, Grenoble, France, 2005

[41] A.V. Vasylyev, P. Weger, W. Bakalski, and W. Simbuerger, “17 GHz 50-60 mW power amplifier in 0.13-um standard CMOS,” IEEE Microwave ad Wireless Components Lett., vol. 16, no. 1, Jan. 2006

[42] A. Komijani and A. Hajimiri, “A 24 GHz, +14.5 dBm fully-integrated power amplifier in 0.18um CMOS,” in Proc. IEEE 2004 Custom Integrated Circuits Conf., Oct. 2004, pp. 561-564.

[43] R. Gilmore and L. Besser, *Practical RF Circuit Design for Modern Wireless Systems: Active Circuits and Systems*, Norwood,MA: Artech House, Inc., 2003

VITA

NAVIN MALLAVARPU

Navin Mallavarpu received his B.S. degree in Electrical Engineering from Rensselaer Polytechnic Institute in Troy, NY and his M.S. degree in Electrical Engineering from Cornell University in Ithaca, NY. He has worked as an intern at Raytheon and Anadigics. At the time of this writing, he is working as a design engineer at Anadigics, Inc. in New Jersey.